

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:05:32 : Search time 234.98 Seconds  
(without alignments)  
250.084 Million cell updates/sec

Title: US-08-956-518A-103  
Sequence: 1 CAGGCGCCGACATGCTGCC.....CAGAGTCTCTGTACATCG 641

Scoring table: IDENTITY\_NUC

Archived: 176461 seqs, 45838279 residues

Issued Patents, NA: \*  
1: /cgn2\_6/ptodata/1/1na/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/1na/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/1na/5C.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/1na/5D.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/1na/PCU9S.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/1na/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	108	16.8	1876	4	US-08-466-589-7	Sequence 7, Appl1
2	40.2	6.3	2990	3	US-08-572-851-1	Sequence 1, Appl1
3	36.6	5.7	1915	3	US-08-466-855A-5	Sequence 5, Appl1
4	36.6	5.7	1915	4	US-08-466-589-11	Sequence 11, Appl1
5	35.4	5.5	1942	4	US-08-604-989A-11	Sequence 11, Appl1
6	35.2	5.5	31571	2	US-08-323-443B-1	Sequence 1, Appl1
7	34.6	5.4	2374	4	US-08-466-589-5	Sequence 5, Appl1
8	34	5.3	3223	4	US-07-980-528-1	Sequence 1, Appl1
9	33.8	5.3	1018	1	US-08-444-083-6	Sequence 6, Appl1
10	33.8	5.3	1018	1	US-08-286-304-6	Sequence 6, Appl1
11	33.8	5.3	1018	1	US-08-442-745-6	Sequence 6, Appl1
12	33.8	5.3	1018	1	US-08-443-129-6	Sequence 6, Appl1
13	33.8	5.3	1018	2	US-08-443-952-6	Sequence 6, Appl1
14	33.8	5.3	1018	2	US-08-443-130-6	Sequence 6, Appl1
15	33.8	5.3	43280	4	US-08-804-227C-1	Sequence 1, Appl1
16	33.8	5.3	1018	5	PCT-US95-04467-6	Sequence 6, Appl1
17	33.6	5.2	8438	1	US-07-945-283-1	Sequence 1, Appl1
18	33.6	5.2	1597	1	US-08-166-316-1	Sequence 1, Appl1
19	33.4	5.2	2064	2	US-08-343-428-1	Sequence 1, Appl1
20	33.4	5.2	1152	5	PCT-US93-06251-51	Sequence 51, Appl1
21	33.2	5.2	388	1	US-07-626-618A-1	Sequence 1, Appl1
22	33.2	5.2	388	1	US-08-333-977-1	Sequence 1, Appl1
23	33.2	5.2	2638	3	US-08-306-691B-46	Sequence 46, Appl1
24	33	5.1	11219	3	US-07-642-734C-1	Sequence 1, Appl1
25	33	5.1	11558	5	PCT-US93-06251-23	Sequence 23, Appl1
26	32.8	5.1	2277	3	US-08-466-855A-1	Sequence 1, Appl1
27	32.6	5.1	1161	1	US-08-086-439C-2	Sequence 2, Appl1
28	32.6	5.1	12001	3	US-08-434-877-2	Sequence 11, Appl1
29	32.6	5.1	4257	4	US-08-458-568A-11	Sequence 1, Appl1
30	32.6	5.1	2923	2	US-08-690-473-1	Sequence 1, Appl1
31	32.4	5.1	2923	4	US-08-377-292-6	Sequence 6, Appl1
32	32.4	5.1	2923	4	US-07-969-847-7	Sequence 7, Appl1
33	32.2	5.0	11613	1	US-08-484-044-10	Sequence 10, Appl1
34	32.2	5.0	1864	1	US-08-353-948-1	Sequence 2, Appl1
35	32	5.0	6961	3	US-08-727-034-2	Sequence 1, Appl1
36	32	5.0	1578	3	US-08-681-129-1	Sequence 1, Appl1
37	32	5.0	3836	4	US-08-216-260-1	Sequence 1, Appl1

38	31.8	5.0	2505	1	US-07-977-434-9	Sequence 9, Appl1
39	31.8	5.0	2504	1	US-08-073-384C-3	Sequence 3, Appl1
40	31.8	5.0	2504	1	US-08-254-359A-3	Sequence 3, Appl1
41	31.8	5.0	2640	1	US-08-384-490-30	Sequence 30, Appl1
42	31.8	5.0	2504	2	US-08-483-043-3	Sequence 3, Appl1
43	31.8	5.0	2640	3	US-08-459-383-30	Sequence 30, Appl1
44	31.8	5.0	2505	3	US-08-458-819-9	Sequence 9, Appl1
45	31.8	5.0	2504	3	US-08-481-238-3	Sequence 3, Appl1

## ALIGNMENTS

RESULT 1  
US-08-466-589-7  
Sequence 7, Application US/08466589  
Patent No. 5837489  
GENERAL INFORMATION:  
APPLICANT: Elliot, Kathryn J.  
APPLICANT: Ellis, Steven B.  
APPLICANT: Harpold, Michael M.  
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE  
RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSPRO Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,589  
FILING DATE: June 5, 1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/028,031  
FILING DATE: March 8, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-9950  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
TELEX:  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1876 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 73..1581  
US-08-466-589-7  
Query Match 16.8%; Score 108; DB 4; Length 1876;  
Best Local Similarity 95.7%; Pred. No. 7.6e-22;  
Matches 11; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 526 CTCGCGATATGCGATGAGCGCTTTGACGCGACATTCACACATTAAGTGTGTAATTC 585  
Db 414 CTATACAGCGCGATGAGCGCTTTGACGCGACATTCACACATTAAGTGTGTAATTC 473  
QY 586 TTCTGGGCAATTGGCAGTACCTGCTCCAGGACATATTCAGAGTTCCTGTACATCG 641





Db	61	CCGGACACCAACATGCGCTGCTCCGCCGGAGGCGTTGGCTGGCGCTGGCGCGGTGCTC	120
QY	441	CTGCACG	447
Db	121	CTGCACG	127

RESULT 2  
US-08-458-568A-11/C  
Sequence 11, Application US/08458568A  
Patent No. 5821339  
GENERAL INFORMATION:  
APPLICANT: Schaffer, Priscilla A.  
APPLICANT: Yeh, Lily  
TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesviruses  
TITLE OF INVENTION: Infections  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:

[illegible]

Accession	Sequence	Position
Db 21.19	GGAGGGGCGAGGGAGCGCGGAGGGGCGAGGGGCGGAGGGGCGAGGGGCGGAGGGGCG	2140
QY 2.23	GGGGCTGTCACGTGAGAGGCGCGCGGGGGGGGGCGGGCGGGGCGCGCCGCGCTC	282
Db 21.19	GAGGGCGGGAGGGGGGCGAGGGGCGGAGGGGGCGAGGGGCGGAGGGGCGGAGGGGCGG	2080
QY 2.83	CTTAAAGGCGCGCGAGGCCGAGCGCGAGGTGCTCTGTGCGCGCAGGCGCGAGGCCCGC	342
Db 20.79	GAGGGGCGAGGGGCGGAGGGGCGGAGGGGCGCGGAGGGGCGG-AGGGGCGGAGGGGCG	2021
QY 3.43	GACAGCGAGAGCTGGAGCGCGCGG 368	
Db 20.20	GAGGGGCGGAGGGGCGGAGGGGCGG 1995	

```

RESULT 3.
US-08-458-568A-11
Sequence 11, Application US/08458568A
Patent No. 5821339
GENERAL INFORMATION:
APPLICANT: Schaffer, Priscilla A.
APPLICANT: Yeh, Lily
TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
TITLE OF INVENTION: Infections
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339r1s
STREET: One Liberty Place, 46th floor
City: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,568A
FILING DATE: 02-JUNE-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/065,146
FILING DATE: 05-MAY-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: leary Ph.D., Kathryn R.
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: DFCI-0029
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 12001 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEetical: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Herpes simplex virus
STRAIN: Herpes Simplex Virus Type 1
US-08-458-568A-11

```

Query Match 11.6% Score 53 DB 3 Length 12001.  
Best Local Similarity 48.7% Pmed No. 0.011.  
Matches 172; Conservative 0; Mismatches 180; Indels 1; Gaps 88  
GGGGGTGAGGTGCCGAGAGCTACCCAGGCGCGGAGTACCTCCCGCTCACACCTCGG 147



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM.nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:05:13 ; Search time 234.98 Seconds  
(Without alignments)  
39.015 Million cell updates/sec

Title: US-08-956-518A-99  
Perfect score: 100  
Sequence: 1 CTCCTTCTACTGCTGATGAG.....CTGCTCCAGGTAAGCTGCA 100

Scoring table: IDENTITY\_NUC

Searched: 176461 seqs, 45838279 residues

abase :

Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/5C.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/5D.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCUS9.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/Backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	83.4	83.4	1876	4	US-08-466-589-7	Sequence 7, App1
2	24.2	24.2	1384	1	US-07-607-538C-1	Sequence 1, App1
3	23.6	23.6	11601	4	US-08-222-617A-3	Sequence 3, App1
4	23.6	23.6	11601	4	US-08-222-617A-24	Sequence 24, App1
5	23	23.0	46899	3	US-08-471-119A-1	Sequence 1, App1
6	22.8	22.8	1810	1	US-08-299-849B-20	Sequence 20, App1
7	22.4	22.4	35100	4	US-08-770-379-17	Sequence 17, App1
8	22.2	22.2	4566	3	US-08-484-101B-37	Sequence 37, App1
9	22	22.0	1447	1	US-08-236-427-2	Sequence 2, App1
10	22	22.0	1444	1	US-08-236-427-9	Sequence 9, App1
11	21.8	21.8	7432	1	US-07-852-260-1	Sequence 1, App1
12	21.8	21.8	1400	1	US-08-464-164-1	Sequence 1, App1
13	21.8	21.8	4046	3	US-08-706-936-1	Sequence 1, App1
14	21.8	21.8	1400	3	US-08-338-057-1	Sequence 1, App1
15	21.8	21.8	7432	4	US-08-461-503-1	Sequence 1, App1
16	21.8	21.8	1400	4	US-08-668-416-1	Sequence 1, App1
17	21.8	21.8	2042	4	US-08-616-844-5	Sequence 5, App1
18	21.8	21.8	2042	4	US-08-589-654-5	Sequence 5, App1
19	21.8	21.8	2022	5	PCT-US96-00996-4	Sequence 4, App1
20	21.6	21.6	1901	3	US-08-153-848-43	Sequence 43, App1
21	21.6	21.6	1225	3	US-08-494-577-11	Sequence 11, App1
22	21.6	21.6	6171	3	US-08-459-568-1	Sequence 1, App1
23	21.6	21.6	6171	4	US-08-399-411-1	Sequence 1, App1
24	21.6	21.6	1238	4	US-08-795-868-11	Sequence 11, App1
25	21.6	21.6	2793	4	US-08-795-868-13	Sequence 13, App1
26	21.6	21.6	1901	4	PCT-US93-11153-43	Sequence 43, App1
27	21.4	21.4	1816	2	US-08-061-889-1	Sequence 1, App1
28	21.4	21.4	1816	3	US-08-462-611-1	Sequence 1, App1
29	21.4	21.4	1915	3	US-08-496-855A-5	Sequence 5, App1
30	21.4	21.4	1224	4	US-08-210-762E-19	Sequence 19, App1
31	21.4	21.4	3351	4	US-08-210-762E-23	Sequence 23, App1
32	21.4	21.4	1915	4	US-08-466-589-11	Sequence 11, App1
33	21.4	21.4	10968	4	US-08-680-327-2	Sequence 2, App1
34	21.4	21.4	1816	5	PCT-US94-05378-1	Sequence 1, App1
35	21.2	21.2	7641	1	US-07-924-028A-5	Sequence 5, App1
36	21	21.0	13254	1	US-08-276-852-156	Sequence 156, App
37	21	21.0	13254	2	US-08-276-852-170	Sequence 170, App

## ALIGNMENTS

c	38	21	21.0	13254	3	US-08-899-575-156	Sequence 156, App
c	39	21	21.0	13254	3	US-08-899-575-170	Sequence 170, App
c	40	21	21.0	13254	3	US-08-899-575-156	Sequence 156, App
c	41	21	21.0	13254	3	US-08-899-575-170	Sequence 170, App
c	42	21	21.0	9126	3	US-08-580-038-26	Sequence 26, App1
c	43	21	21.0	4080	4	US-08-710-249-3	Sequence 3, App1
c	44	21	21.0	1245	4	US-08-750-524-2	Sequence 2, App1
c	45	21	21.0	13254	5	PCT-US95-08743-170	Sequence 170, App

RESULT 1  
US-08-466-589-7  
Sequence 7, Application US/08466589  
Patent No. 5837489  
GENERAL INFORMATION:  
APPLICANT: Eliot, Kathryn J.  
APPLICANT: Ellis, Steven B.  
APPLICANT: Harpold, Michael M.  
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE  
RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & Mcclain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,589  
FILING DATE: June 5, 1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/028,031  
FILING DATE: March 8, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-9950  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
TELEX:  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1876 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 73..1581  
US-08-466-589-7  
Query Match 83.4%; Score 83.4; DB 4; Length 1876;  
Best Local Similarity 98.8%; Pred. No. 2,6e-22;  
Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 9 AGTCCTATAGAGCGCTTGAAGCCACATTCACACACTAGCGTGTGGTGAATTCCTGGG 68  
|||||  
Db 421 AGTCCTATAGAGCGCTTGAAGCCACATTCACACACTAGCGTGTGGTGAATTCCTGGG 480  
QY 69 CATGCCAGTACCTGCTCCAGGTA 93



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:05:17 : Search time 234.98 Seconds  
(without alignments)  
14.435 Million cell updates/sec

Title: US-08-956-518a-100  
Perfect score: 37  
Sequence: 1 ACCCAGCAGCATATTCAGAGTCTGCTACAT G 37

Scoring table: IDENTITY\_NUC

Searched: 176461 seqs, 45838279 residues

- abase :
- 1: Issued\_Patents\_NA:\*
  - 2: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*
  - 3: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*
  - 4: /cgn2\_6/ptodata/1/ina/5C.COMB.seq:\*
  - 5: /cgn2\_6/ptodata/1/ina/5D.COMB.seq:\*
  - 6: /cgn2\_6/ptodata/1/ina/PCITUS9.COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	81.1	1876	4	Sequence 7, Appl1
2	21	56.8	2374	4	Sequence 5, Appl1
3	19.6	53.0	2277	3	Sequence 1, Appl1
4	19.6	53.0	2068	4	Sequence 1, Appl1
5	19.2	51.9	1299	4	Sequence 4, Appl1
6	18.8	50.8	812	5	Sequence 5, Appl1
7	18.2	49.2	1672	1	Sequence 27, Appl1
8	18.2	49.2	1845	1	Sequence 28, Appl1
9	18.2	49.2	1672	1	Sequence 27, Appl1
10	18.2	49.2	1845	1	Sequence 28, Appl1
11	18.2	49.2	1672	1	Sequence 27, Appl1
12	18.2	49.2	1845	1	Sequence 28, Appl1
13	17.8	48.1	2278	1	Sequence 6, Appl1
14	17.8	48.1	2415	1	Sequence 8, Appl1
15	17.8	48.1	2756	1	Sequence 1, Appl1
16	17.8	48.1	1980	2	Sequence 12, Appl1
17	17.8	48.1	7997	3	Sequence 31, Appl1
18	17.8	48.1	7997	3	Sequence 31, Appl1
19	17.8	48.1	264	4	Sequence 7, Appl1
20	17.8	48.1	350	5	Sequence 14, Appl1
21	17.8	48.1	264	5	Sequence 8, Appl1
22	17.4	47.0	5093	2	Sequence 14, Appl1
23	17.4	47.0	5093	2	Sequence 14, Appl1
24	17.4	47.0	3171	5	Sequence 23, Appl1
25	17.2	46.5	1133	1	Sequence 12, Appl1
26	17.2	46.5	2400	1	Sequence 15, Appl1
27	17.2	46.5	2400	1	Sequence 15, Appl1
28	17.2	46.5	2313	1	Sequence 16, Appl1
29	17.2	46.5	100	1	Sequence 17, Appl1
30	17.2	46.5	100	1	Sequence 17, Appl1
31	17.2	46.5	23	1	Sequence 25, Appl1
32	17.2	46.5	23	1	Sequence 25, Appl1
33	17.2	46.5	23	1	Sequence 4, Appl1
34	17.2	46.5	11531	1	Sequence 4, Appl1
35	17.2	46.5	23	1	Sequence 4, Appl1
36	17.2	46.5	11531	2	Sequence 1, Appl1
37	17.2	46.5	11561	3	Sequence 1, Appl1

38	17.2	46.5	1756	4	US-08-466-589-3	Sequence 3, Appl1
39	17.2	46.5	11561	4	US-08-637-640-1	Sequence 1, Appl1
40	17.2	46.5	23	4	US-08-660-963-4	Sequence 4, Appl1
41	17.2	46.5	305	3	PCT-US91-02311-7	Sequence 13, Appl1
42	17	45.9	33	3	US-08-657-192-13	Sequence 5, Appl1
43	17	45.9	531	4	US-08-340-539A-5	Sequence 5, Appl1
44	17	45.9	531	4	US-08-461-532B-5	Sequence 1, Appl1
45	16.8	45.4	17415	5	PCT-US95-07349-1	

ALIGNMENTS

RESULT 1  
US-08-466-589-7  
Sequence 7, Application US/08466589  
Patent No. 5837489  
GENERAL INFORMATION:  
APPLICANT: Eliott, Kathryn J.  
APPLICANT: Ellis, Steven B.  
APPLICANT: Harpold, Michael M.  
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTA Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,589  
FILING DATE: June 5, 1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/028,031  
FILING DATE: March 8, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-9950  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
TELEX:  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1876 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 73..1561  
US-08-466-589-7  
Query Match 81.1% Score 30: DB 4: Length 1876:  
Best Local Similarity 100.0%: Pred. No. 0.00035:  
Matches 30: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 8 CAGGATATTCAGAGTCTGCTACATCG 37  
|||||  
DB 500 CAGGATATTCAGAGTCTGCTACATCG 529

RESULT 2  
US-08-466-589-5  
Sequence 5, Application US/08466589  
Patent No. 5837489  
GENERAL INFORMATION:  
APPLICANT: Eliot, Kathryn J.  
APPLICANT: Ellis, Steven B.  
APPLICANT: Harpold, Michael M.  
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE  
TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,589  
FILING DATE: June 5, 1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/028,031  
FILING DATE: March 8, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-9950  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
TELEX:  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2374 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 184..2067  
08-466-589-5

Query Match 56.8%; Score 21; DB 4; Length 2374;  
Best Local Similarity 73.0%; Pred. No. 2.2;  
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ACCCAACAGGCGATATCAAGAGTTCCTGCTACATCG 37  
DB 637 ACTCCCGCGCATTTACAGAGCTCTGCGACATCG 673

RESULT 3  
US-08-496-855A-1  
Sequence 1, Application US/08496855A  
Patent No. 5801232  
GENERAL INFORMATION:  
APPLICANT: Eliot, Kathryn J.  
APPLICANT: Ellis, Steven B.  
APPLICANT: Harpold, Michael M.  
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE  
TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain

STREET: 1660 Union Street  
CITY: San Diego  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/496,855A  
FILING DATE: 20-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/149,503  
FILING DATE: 08-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/028,031  
FILING DATE: 08-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-9369B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2277 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 166..1755  
US-08-496-855A-1

Query Match 53.0%; Score 19.6; DB 3; Length 2277;  
Best Local Similarity 73.5%; Pred. No. 8.2;  
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 CACACAGCGATATTCAGAGTTCCTGCTACATCG 37  
DB 688 CCCCCGCCATCTACAGAGCTCTGCGACATCG 721

RESULT 4  
US-08-466-589-1  
Sequence 1, Application US/08466589  
Patent No. 5837489  
GENERAL INFORMATION:  
APPLICANT: Eliot, Kathryn J.  
APPLICANT: Ellis, Steven B.  
APPLICANT: Harpold, Michael M.  
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE  
TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,589

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:05:18; See ch time 234.98 Seconds  
(without alignments)  
152.93 Million cell updates/sec

Title: US-08-956-518a-101  
Perfect score: 392  
Sequence: 1 AGACGCAAGGAGGAGGAGTGTG.....TGACGCTCGGAGCTCAAC 392

Scoring table: IDENTITY\_NUC

Searched: 176461 seqs, 45838279 residues

\*base :

1: /cgn2\_6/protdata/1/lna/3A\_COMB.seq:\*  
2: /cgn2\_6/protdata/1/lna/5B\_COMB.seq:\*  
3: /cgn2\_6/protdata/1/lna/5C\_COMB.seq:\*  
4: /cgn2\_6/protdata/1/lna/5D\_COMB.seq:\*  
5: /cgn2\_6/protdata/1/lna/PCITUS9\_COMB.seq:\*  
6: /cgn2\_6/protdata/1/lna/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	72	18.4	1876	4	US-08-466-589-7	Sequence 7, Appl
2	56.4	14.4	12001	3	US-08-458-568A-11	Sequence 11, Appl
3	52.4	13.4	7218	2	US-08-232-463-14	Sequence 14, Appl
4	52.4	13.4	7218	2	US-08-458-568A-3	Sequence 3, Appl
5	52.4	13.4	12001	3	US-08-458-568A-11	Sequence 11, Appl
6	51.8	13.2	4403	4	US-08-284-941-1	Sequence 1, Appl
7	51.8	13.2	4257	4	US-08-690-473-1	Sequence 1, Appl
8	51.8	13.2	4403	5	PCT-US93-02147A-1	Sequence 1, Appl
9	50.2	12.8	1335	5	PCT-US91-06532-1	Sequence 1, Appl
10	49	12.5	4405	1	US-07-885-972A-3	Sequence 3, Appl
11	49	12.5	223	1	US-08-383-761-1	Sequence 1, Appl
12	49	12.5	223	2	US-08-824-277-1	Sequence 1, Appl
13	48.4	12.3	7011	3	US-08-306-691B-42	Sequence 16, Appl
14	48.4	12.3	801	4	US-08-770-379-16	Sequence 4, Appl
15	47.8	12.2	1032	2	US-08-400-422-2	Sequence 2, Appl
16	47.4	12.1	1610	1	US-08-056-051-5	Sequence 5, Appl
17	47.4	12.1	1610	1	US-07-928-611-12	Sequence 12, Appl
18	47.4	12.1	1610	1	US-07-928-611-21	Sequence 21, Appl
19	47.4	12.1	4425	4	US-08-749-169A-1	Sequence 1, Appl
20	47.4	12.1	803	4	US-08-487-811A-12	Sequence 12, Appl
21	47.4	12.1	1610	4	US-08-487-811A-21	Sequence 21, Appl
22	47.4	12.1	803	5	PCT-US93-07370-12	Sequence 12, Appl
23	47.4	12.1	1610	5	PCT-US93-07370-21	Sequence 21, Appl
24	46	11.7	18994	2	US-08-459-586-4	Sequence 4, Appl
25	46	11.7	18994	4	US-08-642-734C-1	Sequence 1, Appl
26	45.8	11.7	11219	3	US-07-642-734C-1	Sequence 4, Appl
27	45.6	11.6	2588	4	US-08-796-414B-6	Sequence 6, Appl
28	44.8	11.4	2362	1	US-08-146-421-4	Sequence 4, Appl
29	44.6	11.4	30001	1	US-08-125-468-1	Sequence 1, Appl
30	44.6	11.4	2823	2	US-08-398-008A-1	Sequence 1, Appl
31	44.6	11.4	1288	3	US-08-440-856A-9	Sequence 9, Appl
32	44.6	11.4	30001	4	US-08-474-833-1	Sequence 1, Appl
33	44.4	11.3	6453	3	US-08-306-691B-14	Sequence 14, Appl
34	44.2	11.3	1723	1	US-07-841-646-28	Sequence 28, Appl
35	44.2	11.3	1723	1	US-07-901-703-10	Sequence 10, Appl
36	44.2	11.3	1723	1	US-08-147-023-28	Sequence 28, Appl
37	44.2	11.3	1723	1	US-08-206-864-3	Sequence 3, Appl

## ALIGNMENTS

C 38	44.2	11.3	1723	2	US-08-278-729A-20	Sequence 20, Appl
C 39	44.2	11.3	1723	2	US-08-480-528A-7	Sequence 7, Appl
C 40	44.2	11.3	1723	2	US-08-479-666-7	Sequence 7, Appl
C 41	44.2	11.3	1723	2	US-08-155-343A-20	Sequence 20, Appl
C 42	44.2	11.3	1723	2	US-08-406-672-20	Sequence 20, Appl
C 43	44.2	11.3	1723	2	US-08-643-563A-20	Sequence 20, Appl
C 44	44.2	11.3	1723	2	US-08-447-570-28	Sequence 28, Appl
C 45	44.2	11.3	1723	5	PCT-US93-10520-7	Sequence 7, Appl

## RESULT 1

US-08-466-589-7  
Sequence 7, Application US/08466589  
Patent No. 5837489  
GENERAL INFORMATION:  
APPLICANT: Elliot, Kathryn J.  
APPLICANT: Ellis, Steven B.  
APPLICANT: Harpold, Michael M.  
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE  
RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & Mcclain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,589  
FILING DATE: June 5, 1995  
CLASSIFICATION: 336  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/028,031  
FILING DATE: March 8, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-9950  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
TELEX:  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1876 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 73..1581  
US-08-466-589-7  
Query Match 18.4%; Score 72; DB 4; Length 1876;  
Best local Similarity 100.0%; Pred. No. 5e-06;  
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 321 GGCCGAGGCGGAGGCGGCGGAGAGCGGAGCGGCGGCTCGCTGAGCT 380  
DB 1 GGCCGAGGCGGAGGCGGCGGCGGAGAGCGGAGCGGCGGCTCGCTGAGCT 60  
QY 381 CCGGAGCTCAAC 392

Db 61 CCGGACTCAAC 72

```

1      RESULT 2
2      US-08-458-568A-11/C
3      : Sequence 11, Application US/08458568A
4      : Patent No. 5621339
5      :
6      : GENERAL INFORMATION:
7      : APPLICANT: Schaffer, Priscilla A.
8      : APPLICANT: Yeh, Lily
9      : TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
10     : TITLE OF INVENTION: Infections
11     : NUMBER OF SEQUENCES: 15
12     :
13     : CORRESPONDENCE ADDRESS:
14     : ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339r1s
15     : STREET: One Liberty Place, 46th floor
16     : CITY: Philadelphia
17     : STATE: PA
18     :
19     : COUNTRY: USA
20     :
21     : ZIP: 19103
22     :
23     : COMPUTER READABLE FORM:
24     : MEDIUM TYPE: Floppy disk
25     : COMPUTER: IBM PC compatible
26     : OPERATING SYSTEM: PC-DOS/MS-DOS
27     : SOFTWARE: Wordperfect 5.1
28     :
29     : CURRENT APPLICATION DATA:
30     : APPLICATION NUMBER: US/08/458,568A
31     : FILING DATE: 02-JUNE-1995
32     : CLASSIFICATION: 435
33     :
34     : PRIOR APPLICATION DATA:
35     : APPLICATION NUMBER: US 08/065,146
36     : FILING DATE: 05-MAY-1993
37     : CLASSIFICATION: 435
38     :
39     : ATTORNEY/AGENT INFORMATION:
40     : NAME: Leary Ph.D., Kathryn R.
41     : REGISTRATION NUMBER: 36,317
42     : REFERENCE/DOCKET NUMBER: PDCI-0029
43     :
44     : TELECOMMUNICATION INFORMATION:
45     : TELEPHONE: (215) 568-3100
46     : TELEFAX: (215) 568-3439
47     :
48     : INFORMATION FOR SEQ ID NO: 11:
49     : SEQUENCE CHARACTERISTICS:
50     : LENGTH: 12001 base pairs
51     : TYPE: nucleic acid
52     : STRANDEDNESS: double
53     : TOPOLOGY: linear
54     :
55     : MOLECULE TYPE: DNA (genomic)
56     :
57     : HYPOTHETICAL: NO
58     : ANTI-SENSE: NO
59     :
60     : ORIGINAL SOURCE:
61     : ORGANISM: Herpes simplex virus
62     : STRAIN: Herpes Simplex Virus Type 1
63     :
64     : US-08-458-568A-11

```

	Query Match	14.4%	Score 56.4	DB 3	Length 12001
	Best Local Similarity	50.3%	Pred. NO. 0.002		
Matches	164	Conservative	0	Mismatches 111	Indels 1; Gaps 1
QY	43	GGCCTGCGCCAGAGCGCCGACGCCGAGAGCCCCCTTGATGA3ACTGGGGGTGAGCTGCC	102		
Db	2319	GGGAGGGGGCGAAGGGCGGGAGGGGGCGAGGGCGCGAGAGGGCGGAGGGCGGGAGGGGG	2260		
QY	103	CGAGAGCTAACCAAGCCCGSGAGTAACTCCGCTAACACTTCGGGCTTCAGATTCCCTGGG	162		
Db	2259	CGAAGGGCGGGAGGGGGCGAGGGGGCGGGAGGGGGCGAGGGGGCGAGAGGGGGCGAGAGGGCG	2200		
QY	163	TGGCCGCGCAGAGAGCTGTGCCCGGGCTGAGAGGATGCGCGGCGGGGGAACGGGGGGGGGG	222		
Db	2199	GGAAGGGGGCGAAGGGGGCGGAGGGGGCGAGAGGGCGGAGAGGGCGGACAGGGGGCGGAGAAGGGGG	2140		
QY	223	GGGGCTTGTCACGTGAGAGAGCGCGCGGGGGCGGGCGGGGGGGCGGCGCGCGCCCGGCTC	282		

[illegible]

```

1      RESULT      3
2      US-08-232-463-14/c
3      : Sequence 14, Application US/08232463
4      : Patent No. 5670367
5      : GENERAL INFORMATION:
6      : APPLICANT: DORNER, F.
7      : APPLICANT: SCHEFFLINGER, F.
8      : APPLICANT: FALKNER, F. G.
9      : TITLE OF INVENTION: RECOMBINANT FOXP2 VIRUS
10     : NUMBER OF SEQUENCES: 52
11     : CORRESPONDENCE ADDRESS:
12     : ADDRESSEE: Foley & Lardner
13     : STREET: 1800 Diagonal Road, Suite 500
14     : City: Alexandria
15     : STATE: VA
16     : COUNTRY: USA
17     : ZIP: 22313-0299
18     : COMPUTER READABLE FORM:
19     : MEDIUM TYPE: Floppy disk
20     : COMPUTER: IBM PC compatible
21     : OPERATING SYSTEM: PC-DOS/MS-DOS
22     : SOFTWARE: PatentIn Release #1.0, Version #1.25
23     : CURRENT APPLICATION DATA:
24     : APPLICATION NUMBER: US/08/232.463
25     : FILING DATE:
26     : CLASSIFICATION: 435
27     : PRIOR APPLICATION DATA:
28     : APPLICATION NUMBER: US/07/935.313
29     : FILING DATE:
30     : APPLICATION NUMBER: EP 91 114 300.6
31     : FILING DATE: 26-AUG-1991
32     : ATTORNEY/AGENT INFORMATION:
33     : NAME: BENT, Stephen A.
34     : REGISTRATION NUMBER: 29,768
35     : REFERENCE/DOCKET NUMBER: 30472/114 IMMU
36     : TELECOMMUNICATION INFORMATION:
37     : TELEPHONE: (703)836-9300
38     : TELEFAX: (703)683-4109
39     : TELEX: 899149
40     : INFORMATION FOR SEQ. ID NO.: 14:
41     : SEQUENCE CHARACTERISTICS:
42     : LENGTH: 7218 base pairs
43     : TYPE: nucleic acid
44     : STRANDEDNESS: single
45     : TOPOLOGY: linear
46     : IMMEDIATE SOURCE:
47     : CLONE: PT29pc-F15
48     :
49     :
50     :
51     :
52     :
53     :
54     :
55     :
56     :
57     :
58     :
59     :
60     :
61     :
62     :
63     :
64     :
65     :
66     :
67     :
68     :
69     :
70     :
71     :
72     :
73     :
74     :
75     :
76     :
77     :
78     :
79     :
80     :
81     :
82     :
83     :
84     :
85     :
86     :
87     :
88     :
89     :
90     :
91     :
92     :
93     :
94     :
95     :
96     :
97     :
98     :
99     :
100    :
101    :
102    :
103    :
104    :
105    :
106    :
107    :
108    :
109    :
110    :
111    :
112    :
113    :
114    :
115    :
116    :
117    :
118    :
119    :
120    :
121    :
122    :
123    :
124    :
125    :
126    :
127    :
128    :
129    :
130    :
131    :
132    :
133    :
134    :
135    :
136    :
137    :
138    :
139    :
140    :
141    :
142    :
143    :
144    :
145    :
146    :
147    :
148    :
149    :
150    :
151    :
152    :
153    :
154    :
155    :
156    :
157    :
158    :
159    :
160    :
161    :
162    :
163    :
164    :
165    :
166    :
167    :
168    :
169    :
170    :
171    :
172    :
173    :
174    :
175    :
176    :
177    :
178    :
179    :
180    :
181    :
182    :
183    :
184    :
185    :
186    :
187    :
188    :
189    :
190    :
191    :
192    :
193    :
194    :
195    :
196    :
197    :
198    :
199    :
200    :
201    :
202    :
203    :
204    :
205    :
206    :
207    :
208    :
209    :
210    :
211    :
212    :
213    :
214    :
215    :
216    :
217    :
218    :
219    :
220    :
221    :
222    :
223    :
224    :
225    :
226    :
227    :
228    :
229    :
230    :
231    :
232    :
233    :
234    :
235    :
236    :
237    :
238    :
239    :
240    :
241    :
242    :
243    :
244    :
245    :
246    :
247    :
248    :
249    :
250    :
251    :
252    :
253    :
254    :
255    :
256    :
257    :
258    :
259    :
260    :
261    :
262    :
263    :
264    :
265    :
266    :
267    :
268    :
269    :
270    :
271    :
272    :
273    :
274    :
275    :
276    :
277    :
278    :
279    :
280    :
281    :
282    :
283    :
284    :
285    :
286    :
287    :
288    :
289    :
290    :
291    :
292    :
293    :
294    :
295    :
296    :
297    :
298    :
299    :
300    :
301    :
302    :
303    :
304    :
305    :
306    :
307    :
308    :
309    :
310    :
311    :
312    :
313    :
314    :
315    :
316    :
317    :
318    :
319    :
320    :
321    :
322    :
323    :
324    :
325    :
326    :
327    :
328    :
329    :
330    :
331    :
332    :
333    :
334    :
335    :
336    :
337    :
338    :
339    :
340    :
341    :
342    :
343    :
344    :
345    :
346    :
347    :
348    :
349    :
350    :
351    :
352    :
353    :
354    :
355    :
356    :
357    :
358    :
359    :
360    :
361    :
362    :
363    :
364    :
365    :
366    :
367    :
368    :
369    :
370    :
371    :
372    :
373    :
374    :
375    :
376    :
377    :
378    :
379    :
380    :
381    :
382    :
383    :
384    :
385    :
386    :
387    :
388    :
389    :
390    :
391    :
392    :
393    :
394    :
395    :
396    :
397    :
398    :
399    :
400    :
401    :
402    :
403    :
404    :
405    :
406    :
407    :
408    :
409    :
410    :
411    :
412    :
413    :
414    :
415    :
416    :
417    :
418    :
419    :
420    :
421    :
422    :
423    :
424    :
425    :
426    :
427    :
428    :
429    :
430    :
431    :
432    :
433    :
434    :
435    :
436    :
437    :
438    :
439    :
440    :
441    :
442    :
443    :
444    :
445    :
446    :
447    :
448    :
449    :
450    :
451    :
452    :
453    :
454    :
455    :
456    :
457    :
458    :
459    :
460    :
461    :
462    :
463    :
464    :
465    :
466    :
467    :
468    :
469    :
470    :
471    :
472    :
473    :
474    :
475    :
476    :
477    :
478    :
479    :
480    :
481    :
482    :
483    :
484    :
485    :
486    :
487    :
488    :
489    :
490    :
491    :
492    :
493    :
494    :
495    :
496    :
497    :
498    :
499    :
500    :
501    :
502    :
503    :
504    :
505    :
506    :
507    :
508    :
509    :
510    :
511    :
512    :
513    :
514    :
515    :
516    :
517    :
518    :
519    :
520    :
521    :
522    :
523    :
524    :
525    :
526    :
527    :
528    :
529    :
530    :
531    :
532    :
533    :
534    :
535    :
536    :
537    :
538    :
539    :
540    :
541    :
542    :
543    :
544    :
545    :
546    :
547    :
548    :
549    :
550    :
551    :
552    :
553    :
554    :
555    :
556    :
557    :
558    :
559    :
560    :
561    :
562    :
563    :
564    :
565    :
566    :
567    :
568    :
569    :
570    :
571    :
572    :
573    :
574    :
575    :
576    :
577    :
578    :
579    :
580    :
581    :
582    :
583    :
584    :
585    :
586    :
587    :
588    :
589    :
590    :
591    :
592    :
593    :
594    :
595    :
596    :
597    :
598    :
599    :
600    :
601    :
602    :
603    :
604    :
605    :
606    :
607    :
608    :
609    :
610    :
611    :
612    :
613    :
614    :
615    :
616    :
617    :
618    :
619    :
620    :
621    :
622    :
623    :
624    :
625    :
626    :
627    :

```

Query Match	13.4%	Score 52.4	DB 2	Length 7218
Best Local Similarity	3.6%	Pred. No. 0.014	Mismatches 133	Indels 0
Matches 14	Conservative 21	Mismatches 133	Indels 0	Gaps 0
QY 1	AGACGCGAAGGAGAGGTAGACCTGCGCTTGGGCGACCCCTGGCTGGCCAGAGGCGG	60		
Db 1426	RR	1367		
QY 61	AGGCGGAGAGCCCGCTGGTGAGACTGGGGGTGAGAGTGGCCGAGCGTACCAGCGCC	120		
Db 1366	RR	1307		

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:03:28 ; Search time 234.98 Seconds  
(without alignments)  
268.811 Million cell updates/sec

Title: US-08-956-518A-102  
Perfect score: 689  
Sequence: 1 AGCCCTTCCAGCGGTAG.....CAGAGTCTCTGCTACATCG 689

Scoring table: IDENTITY\_NDC

Archived: 176461 seqs, 45838279 residues

base :  
Issued\_Patents\_NA.\*  
1: /cgn2\_5/ptoddata/1/lna/5A.COMB.seq.\*  
2: /cgn2\_6/ptoddata/1/lna/5B.COMB.seq.\*  
3: /cgn2\_6/ptoddata/1/lna/5C.COMB.seq.\*  
4: /cgn2\_6/ptoddata/1/lna/5D.COMB.seq.\*  
5: /cgn2\_6/ptoddata/1/lna/PTUS9.COMB.seq.\*  
6: /cgn2\_6/ptoddata/1/lna/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	108	15.7	1876	4	US-08-466-589-7
2	40.2	5.8	2990	3	US-08-572-951-1
3	38	5.5	1578	3	US-08-681-129-1
4	36.6	5.3	3223	1	US-07-980-528-1
5	36.6	5.3	1915	3	US-08-496-855A-5
6	35.6	5.3	1915	4	US-08-466-589-11
7	35.4	5.1	1942	4	US-08-604-989A-11
8	35.2	5.1	31571	2	US-08-323-443B-1
9	34.8	5.1	12001	4	US-08-458-568A-11
10	34.6	5.0	2374	4	US-08-466-589-5
11	33.8	4.9	1018	1	US-08-444-083-6
12	33.8	4.9	1018	1	US-08-286-304-6
13	33.8	4.9	1018	1	US-08-442-745-6
14	33.8	4.9	1018	1	US-08-443-129-6
15	33.8	4.9	1018	2	US-08-443-952-6
16	33.8	4.9	1018	2	US-08-443-130-6
17	33.8	4.9	43280	2	US-08-804-227C-1
18	33.8	4.9	1018	5	PCT-US95-04467-6
19	33.6	4.9	8438	1	US-07-945-283-1
20	33.6	4.9	3068	1	US-07-984-044A-3
21	33.6	4.9	1597	1	US-08-166-316-1
22	33.6	4.9	3068	2	US-08-458-393-3
23	33.4	4.8	2064	2	US-08-343-428-1
24	33.4	4.8	1192	5	PCT-US93-06251-51
25	33.2	4.8	388	1	US-07-626-618A-1
26	33.2	4.8	388	1	US-08-333-977-1
27	33.2	4.8	2638	3	US-08-306-691B-46
28	33.2	4.8	4257	4	US-07-642-734C-1
29	33	4.8	11219	4	US-08-690-473-1
30	33	4.8	11558	5	PCT-US93-06251-23
31	32.8	4.8	2277	3	US-08-496-855A-2
32	32.6	4.7	1161	1	US-08-086-439C-2
33	32.6	4.7	1161	1	US-08-434-877-2
34	32.6	4.7	12001	3	US-08-458-568A-11
35	32.6	4.7	4257	4	US-08-690-473-1
36	32.4	4.7	2923	2	US-08-377-292-6
37	32.4	4.7	2923	4	US-07-989-847-7

## ALIGNMENTS

C 38	32.2	4.7	11613	1	US-08-484-044-10	Sequence 10, Appl
C 39	32.2	4.7	1864	1	US-08-353-948-1	Sequence 1, Appl
C 40	32	4.6	6961	2	US-08-727-034-2	Sequence 2, Appl
C 41	32	4.6	3836	4	US-08-216-260-1	Sequence 1, Appl
C 42	31.8	4.6	2505	1	US-07-977-434-9	Sequence 9, Appl
C 43	31.8	4.6	2504	1	US-08-073-384C-3	Sequence 3, Appl
C 44	31.8	4.6	3070	1	US-08-428-732-3	Sequence 3, Appl
C 45	31.8	4.6	2504	1	US-08-254-359A-3	Sequence 3, Appl

```
RESULT 1
US-08-466-589-7
; Sequence 7, Application US/08466589
; Patent No. 5837489
; GENERAL INFORMATION:
; APPLICANT: Eliot, Kathryn J.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Harpold, Michael M.
; TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
; TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,589
; FILING DATE: June 5, 1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/028,031
; FILING DATE: March 8, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9950
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1876 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 73..1581
; US-08-466-589-7

Query Match 15.7%; Score 108; DB 4; Length 1876;
Best Local Similarity 95.7%; Pred. No. 2.5e-21;
Matches 111; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 574 CCGCGATATGCTGATGAGCGGCTTGAGCCGACATTCACACACTACGCTGCTAATTC 633
DB 414 CATACACTGCTGATGAGCGGCTTGAGCCGACATTCACACACTACGCTGCTAATTC 473
QY 634 TTCTGGGATTCGCCAGTACTGCTCCAGGATATTCAGAGTCTCTGCTACATCG 689
```

Db 474 TTCTGGGCGATTGCCAGTACTGCTCCCTCCAGGCGATTTCAGAGTTCCTGCTACTGCG 529

RESULT 2  
US-08-572-951-1  
Sequence 1, Application US/08572951  
Patent No. 5824790

GENERAL INFORMATION:  
APPLICANT: KEELING, PETER L.  
APPLICANT: KNIGHT, MARY E.  
APPLICANT: GUAN, HANPING  
TITLE OF INVENTION: MODIFICATION OF STARCH  
TITLE OF INVENTION: SYNTHESIS IN PLANTS  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DABY & CUSHMAN  
ADDRESSEE: Intellectual Property Group of  
STREET: Pillsbury Madison & Sutro LLP  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3918

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/572,951  
FILING DATE: 15-DEC-1995  
CLASSIFICATION: 800

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/346,602  
FILING DATE: 29-NOV-1994  
CLASSIFICATION: 800

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/263,921  
FILING DATE: 21-JUN-1994  
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:  
NAME: Paul N. Kokulis

REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: 222957/1.02.15C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944

AFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2990 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA

US-08-572-951-1

Query Match 5.8%; Score 40.2; DB 3; Length 2990;  
Best Local Similarity 57.6%; Pred. N3.0.025; Mismatches 53; Indels 0; Gaps 0;  
Matches 72; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 170 GCTCCACGAGGATGACGCGCGGAGAGTTGAGACCGCGAGAGCTCGCGCGGAGGCGC 229  
Db 693 GCTCCACGAGGATGACGCGCGGAGAGTTGAGACCGCGAGAGCTCGCGCGGAGGCGC 752  
QY 230 CGCGCGGATGACGCGCGGAGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 289  
Db 753 GCGCGCGGATGACGCGCGGAGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 812  
QY 290 CGCGCG 294  
Db 813 GCGCG 817

RESULT 3

US-08-681-129-1  
Sequence 1, Application US/08681129  
Patent No. 5738654

GENERAL INFORMATION:  
APPLICANT: Mettenleiter, Thomas Cristoph  
TITLE OF INVENTION: Pseudorabies virus vaccine  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Organon Teknika Corporation  
STREET: 1330 Piccard Drive  
CITY: Rockville  
STATE: Maryland  
COUNTRY: U.S.A.  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/681,129  
FILING DATE:  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/244,446  
FILING DATE: 02-JUN-1994  
CLASSIFICATION: 424

APPLICATION NUMBER: EP 92.203.079.6  
FILING DATE: 06-OCT-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Gornley, Mary E.

REGISTRATION NUMBER: 34,409  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 258-5200

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1578 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Pseudorabies virus  
STRAIN: Kaplan

US-08-681-129-1

Query Match 5.5%; Score 38; DB 3; Length 1578;  
Best Local Similarity 50.5%; Pred. N0.0.08; Mismatches 90; Indels 0; Gaps 0;  
Matches 92; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 53 GCGGCTTGACGCGGAGCGCGGCTCTGCTGAGTGTGATTAAGAGGAGGAGGAGGAGGAG 112  
Db 1019 GCGGCTTGACGCGGAGCGCGGCTCTGCTGAGTGTGATTAAGAGGAGGAGGAGGAGGAG 1078  
QY 113 CAGGCGGAGGATGACGCGCGGAGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 172  
Db 1079 GCGGCGGAGGATGACGCGCGGAGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1138  
QY 173 CCGAGGAGGATGACGCGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 232  
Db 1139 CCGAGGAGGATGACGCGCGGAGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1198  
QY 233 CC 234  
Db 1199 CC 1200

RESULT 4  
US-07-980-528-1/c





NI	9990310
XX	
DT	22-SEP-1995 (Rel. 45, Created)
DT	22-SEP-1995 (Rel. 45, Last updated, Version 1)
XX	
DE	y025gf08.r1 Homo sapiens cDNA 5' similar to contains MER22
XX	repetitive element ;.
XX	
KW	EST.
OS	Homo sapiens (human)
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
CC	Primates; Catarrhini; Hominoidea; Homo.
XX	
RA	[1]
RP	1-437
RA	Hillier L., Clark N., Dubugue T., Elliston K., Hawkins M., Holman M.,
RA	Hultman M., Kucaba T., Le M., Lennon G., Maier M., Parsons J.,
RA	Rifkin L., Rohlfing T., Soares W., Tan F., Trevaskis E., Waterston R.,
RA	Williamson A., Wohlmann P., Wilson R.;
RA	"The WashU-Merck EST Project";
RA	Unpublished.
N-	Contact: Wilson RK WashU-Merck EST Project Washington University
CC	School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,
CC	MO 63108 Tel.: 314 286 1800 fax: 314 286 1810 Email:
CC	est@wustl.wustl.edu High quality sequence stops: 313 Source: IMAGE
CC	Consortium, LNL This clone is available royalty-free through LNL
CC	; contact the IMAGE Consortium (info@image.llnl.gov) for further
CC	Information. NCBI gi: 990310
XX	
XX	
XX	
FT	Key
FH	Location/Qualifiers
FT	source
FT	1..437
FT	/db_xref="taxon:9606"
FT	/organism="Homo sapiens"
FT	<1..>437
XX	mRNA
SO	Sequence 437 bp; 67 A; 128 C; 127 G; 107 T; 8 other;
Query Match	12.7%; Score 7; DB 18; Length 437;
Best Local Similarity	100.0%; Pred No. 5.5e+04;
Matches	7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	49 CTCGACG 55
--	
--	328 CTCGACG 334
RESULT	2
D11490	
LOCUS	D11490 480 bp mRNA EST 02-DEC-1992.
DEFINITION	H00M000AA22 Liver HepG2 cell line. Homo sapiens cDNA clone a-22.
ACCESSION	D11490
NID	92148209
VERSION	D11490.1 GI:2148209
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Primates; Catarrhini; Hominoidea; Homo.
TITLE	Okubo,K., Horii,N., Matoba,R., Miyama,T., Fukushima,A., Kojima,Y.
JOURNAL	Large scale cDNA sequencing for analysis of quantitative and
MEDLINE	qualitative aspects of gene expression
COMMENT	Nature Genet. 2, 173-179 (1992)
COMMENT	94258199
CONTACT:	Contact: Kouzaku Okubo, Naohiro Horii, Ryo Matoba, Toshiyuki
CONTACT:	Miyama, Atsushi Fukushima, Yuko Kojima & Kenichi Matsubara

```

Institute for Molecular and Cellular Biology
Osaka University
1-3 Yamada-oka, Suita, Osaka 565, Japan.
Location/Qualifiers
1. .480
/organism="Homo sapiens"
/db_xref="GDB:DG57502E"
/db_xref="taxon:9606"
/clone="a-22"
/clone_id="Liver Hepg2 cell line."
/lab_host="E.coli"
/note="3'-directed regional cDNA library. Cleaved by MboI
and transformed into E.coli."
BASE COUNT 106 a 114 c 121 g 77 t 62 others
ORIGIN

Query Match 12.7%; Score 7; DB 20; Length 480;
Best Local Similarity 100.0%; Pred. No. 5.5e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 CTCGACG 55
|||||
Db 137 CTCGACG 143

RESULT 3
D11514/c 383 bp mRNA EST 02-DEC-1992
LOCUS HUM00000a20 Liver Hepg2 cell line. Homo sapiens cDNA clone a20, mRNA
DEFINITION sequence.
ACCESSION D11514
NID G2148192
VERSION D11514.1
KEYWORDS EST.
SOURCE GI:2148192
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 383)
Okubo,K., Hori,N., Matoba,R., Niijima,T., Fukushima,A., Kojima,Y.
and Matsubara,K.
Large scale cDNA sequencing for analysis of quantitative and
qualitative aspects of gene expression
Nature Genet. 2, 173-179 (1992)
94258199

COMMENT
JOURNAL Contact: Kousaku Okubo, Naohiro Hori, Ryo Matoba, Toshiyuki
MEDLINE Niijima, Atsushi Fukushima, Yuko Kojima & Kenichi Matsubara
Institute for Molecular and Cellular Biology
Osaka University
1-3 Yamada-oka, Suita, Osaka 565, Japan.
Location/Qualifiers
1. .383
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="a20"
/clone_id="Liver Hepg2 cell line."
/lab_host="E.coli"
/note="3'-directed regional cDNA library. Cleaved by MboI
and transformed into E.coli."
BASE COUNT 106 a 106 c 104 g 59 t 8 others
ORIGIN

FEATURES
source

Query Match 12.7%; Score 7; DB 20; Length 383;
Best Local Similarity 100.0%; Pred. No. 5.4e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 CTCGACG 55
|||||
Db 51 CTCGACG 45

```

RESULT 4  
 D11560 532 bp mRNA EST 02-DEC-1992  
 LOCUS HUM0C12D08 Liver HepG2 cell line. Homo sapiens cDNA clone c12d08,  
 DEFINITION mRNA sequence.  
 ACCESSION D11560  
 NID 92148706  
 VERSION D11560.1 GI:2148706  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 532)  
 Okubo,K., Hori,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y.  
 and Matsubara,K.  
 Large scale cDNA sequencing for analysis of quantitative and  
 qualitative aspects of gene expression  
 Nature Genet. 2, 173-179 (1992)  
 94258199

COMMENT  
 JOURNAL MEDLINE  
 CONTACT: Kousaku Okubo, Naohiro Hori, Ryo Matoba, Toshiyuki  
 Niiyama, Atsushi Fukushima, Yuko Kojima & Kenichi Matsubara  
 Institute for Molecular and Cellular Biology  
 Osaka University  
 1-3 Yamada-oka, Suita, Osaka 565, Japan.  
 Location/Qualifiers  
 1..532  
 /organism="Homo sapiens"  
 /db\_xref="GDB:D05862E"  
 /db\_xref="taxon:9606"  
 /clone="c12d08"  
 /clone\_1lb="Liver HepG2 cell line."  
 /lab\_host="E.coli"  
 /note="3'-directed regional cDNA library. Cleaved by MboI  
 and transformed into E.coli."  
 and transformed into E.coli." 43 others

BASE COUNT 124 a 112 c 109 g 144 t  
 ORIGIN

Query Match 12.7%; Score 7; DB 20; Length 532;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 CTGCACG 55  
 |||||  
 247 CTGCACG 253

RESULT 5  
 D11631 376 bp mRNA EST 02-DEC-1992  
 LOCUS HUM00C202 Liver HepG2 cell line. Homo sapiens cDNA clone c202,  
 DEFINITION mRNA sequence.  
 ACCESSION D11631  
 NID 92148223  
 VERSION D11631.1 GI:2148223  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 376)  
 Okubo,K., Hori,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y.  
 and Matsubara,K.  
 Large scale cDNA sequencing for analysis of quantitative and  
 qualitative aspects of gene expression  
 Nature Genet. 2, 173-179 (1992)  
 94258199

COMMENT  
 JOURNAL MEDLINE  
 CONTACT: Kousaku Okubo, Naohiro Hori, Ryo Matoba, Toshiyuki  
 Niiyama, Atsushi Fukushima, Yuko Kojima & Kenichi Matsubara

Institute for Molecular and Cellular Biology  
 Osaka University  
 1-3 Yamada-oka, Suita, Osaka 565, Japan.  
 Location/Qualifiers  
 1..376  
 /organism="Homo sapiens"  
 /db\_xref="GDB:D057516E"  
 /db\_xref="taxon:9606"  
 /clone="c202"  
 /clone\_1lb="Liver HepG2 cell line."  
 /lab\_host="E.coli"  
 /note="3'-directed regional cDNA library. Cleaved by MboI  
 and transformed into E.coli."  
 and transformed into E.coli." 57 t

BASE COUNT 107 a 107 c 105 g 57 t  
 ORIGIN

Query Match 12.7%; Score 7; DB 20; Length 376;  
 Best Local Similarity 100.0%; Pred. No. 5.3e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 CTGCACG 55  
 |||||  
 Db 51 CTGCACG 45

RESULT 6  
 D11791 411 bp mRNA EST 02-DEC-1992  
 LOCUS HUMHM01G02 Liver HepG2 cell line. Homo sapiens cDNA clone hm01g02,  
 DEFINITION mRNA sequence.  
 ACCESSION D11791  
 NID 92155069  
 VERSION D11791.1 GI:2155069  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 411)  
 Okubo,K., Hori,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y.  
 and Matsubara,K.  
 Large scale cDNA sequencing for analysis of quantitative and  
 qualitative aspects of gene expression  
 Nature Genet. 2, 173-179 (1992)  
 94258199

COMMENT  
 JOURNAL MEDLINE  
 CONTACT: Kousaku Okubo, Naohiro Hori, Ryo Matoba, Toshiyuki  
 Niiyama, Atsushi Fukushima, Yuko Kojima & Kenichi Matsubara  
 Institute for Molecular and Cellular Biology  
 Osaka University  
 1-3 Yamada-oka, Suita, Osaka 565, Japan.  
 Location/Qualifiers  
 1..411  
 /organism="Homo sapiens"  
 /db\_xref="GDB:D058337E"  
 /db\_xref="taxon:9606"  
 /clone="hm01g02"  
 /clone\_1lb="Liver HepG2 cell line."  
 /lab\_host="E.coli"  
 /note="3'-directed regional cDNA library. Cleaved by MboI  
 and transformed into E.coli."  
 and transformed into E.coli." 2 others

BASE COUNT 116 a 119 c 109 g 65 t  
 ORIGIN

Query Match 12.7%; Score 7; DB 20; Length 411;  
 Best Local Similarity 100.0%; Pred. No. 5.4e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 CTGCACG 55  
 |||||  
 Db 87 CTGCACG 81

**RESULT 7**  
**LOCUS** D11889/c 404 bp mRNA EST 02-DEC-1992  
**DEFINITION** HUMHMO5E09 Liver Hepg2 cell line. Homo sapiens CDNA clone hm05e09, mRNA sequence.  
**ACCESSION** D11889  
**NID** G2155149  
**VERSION** D11889.1 GI:2155149  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
**AUTHORS** Okubo,K., Horii,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y. and Matsubara,K.  
**TITLE** Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression  
**JRNAL** Nature Genet. 2, 173-179 (1992)  
**MEDLINE** 94258199  
**COMMENT**  
**FEATURES**  
**source**  
 Contact: Kousaku Okubo, Naohiro Horii, Ryo Matoba, Toshiyuki Niiyama, Atsushi Fukushima, Yoko Kojima & Kenichi Matsubara  
 Institute for Molecular and Cellular Biology  
 Osaka University  
 1-3 Yamada-oka, Suita, Osaka 565, Japan.  
 Location/Qualifiers  
 1..404  
 /organism="Homo sapiens"  
 /db\_xref="GDB:D058435E"  
 /db\_xref="taxon:9606"  
 /clone="hm05e09"  
 /clone\_1lb="Liver Hepg2 cell line."  
 /lab\_host="E.coli"  
 /note="3'-directed regional cDNA library. Cleaved by MboI and transformed into E.coli."  
**BASE COUNT** 94 a 125 c 96 g 84 t 5 others  
**ORIGIN**  
 Query Match 12.7%; Score 7; DB 20; Length 404;  
 Best Local Similarity 100.0%; Pred. No. 5.4e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
**OV** 49 CTCGACG 55  
 |||||  
 332 CTCGACG 326  
**RESULT 8**  
**LOCUS** D11893/c 244 bp mRNA EST 02-DEC-1992  
**DEFINITION** HUMHMO5F08 Liver Hepg2 cell line. Homo sapiens CDNA clone hm05f08, mRNA sequence.  
**ACCESSION** D11893  
**NID** G2155151  
**VERSION** D11893.1 GI:2155151  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
**AUTHORS** Okubo,K., Horii,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y. and Matsubara,K.  
**TITLE** Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression  
**JOURNAL** Nature Genet. 2, 173-179 (1992)  
**MEDLINE** 94258199  
**COMMENT**  
 Contact: Kousaku Okubo, Naohiro Horii, Ryo Matoba, Toshiyuki

**FEATURES**  
**source**  
 Niiyama, Atsushi Fukushima, Yoko Kojima & Kenichi Matsubara  
 Institute for Molecular and Cellular Biology  
 Osaka University  
 1-3 Yamada-oka, Suita, Osaka 565, Japan.  
 Location/Qualifiers  
 1..244  
 /organism="Homo sapiens"  
 /db\_xref="GDB:D058435E"  
 /db\_xref="taxon:9606"  
 /clone="hm05f08"  
 /clone\_1lb="Liver Hepg2 cell line."  
 /lab\_host="E.coli"  
 /note="3'-directed regional cDNA library. Cleaved by MboI and transformed into E.coli."  
**BASE COUNT** 69 a 72 c 49 g 51 t 3 others  
**ORIGIN**  
 Query Match 12.7%; Score 7; DB 20; Length 244;  
 Best Local Similarity 100.0%; Pred. No. 5e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
**OV** 49 CTCGACG 55  
 |||||  
 171 CTCGACG 165  
**RESULT 9**  
**LOCUS** D12042/c 458 bp mRNA EST 02-DEC-1992  
**DEFINITION** HUM0005155 Liver Hepg2 cell line. Homo sapiens CDNA clone sl55, mRNA sequence.  
**ACCESSION** D12042  
**NID** G2148299  
**VERSION** D12042.1 GI:2148299  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
**AUTHORS** Okubo,K., Horii,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y. and Matsubara,K.  
**TITLE** Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression  
**JOURNAL** Nature Genet. 2, 173-179 (1992)  
**MEDLINE** 94258199  
**COMMENT**  
 Contact: Kousaku Okubo, Naohiro Horii, Ryo Matoba, Toshiyuki Niiyama, Atsushi Fukushima, Yoko Kojima & Kenichi Matsubara  
 Institute for Molecular and Cellular Biology  
 Osaka University  
 1-3 Yamada-oka, Suita, Osaka 565, Japan.  
 Location/Qualifiers  
 1..458  
 /organism="Homo sapiens"  
 /db\_xref="GDB:D057600E"  
 /db\_xref="taxon:9606"  
 /clone="sl55"  
 /clone\_1lb="Liver Hepg2 cell line."  
 /lab\_host="E.coli"  
 /note="3'-directed regional cDNA library. Cleaved by MboI and transformed into E.coli."  
**BASE COUNT** 134 a 105 c 109 g 101 t 9 others  
**ORIGIN**  
 Query Match 12.7%; Score 7; DB 20; Length 458;  
 Best Local Similarity 100.0%; Pred. No. 5.5e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
**OV** 49 CTCGACG 55  
 |||||

Db 295 CTCGACG 289

RESULT 10  
LOCUS D12100/c 397 bp mRNA EST 07-OCT-1996  
DEFINITION HUM0005173 Liver HepG2 cell line. Homo sapiens cDNA clone sl73 3',  
RNA sequence.

ACCESSION D12100  
NID 9767711  
VERSION D12100.1 GI:767711  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y.  
and Matsubara, K.  
TITLE Large scale cDNA sequencing for analysis of quantitative and  
qualitative aspects of gene expression  
NATURE GENET. 2, 173-179 (1992)

JOURNAL MEDLINE 94258199  
COMMENT 2 (bases 1 to 397)  
Murakawa, K., Matsubara, K., Fukushima, A., Yoshii, J. and Okubo, K.  
Chromosomal assignments of 3'-directed partial cDNA sequences  
representing novel genes expressed in granulocytoid cells  
GENOMICS 23, 379-389 (1994)  
95137584

FEATURES  
SOURCE Contact: Kouzaku Okubo, Naohiro Hori, Ryo Matoba, Toshiyuki  
Niiyama, Atsushi Fukushima, Yoko Kojima & Kenichi Matsubara  
Institute for Molecular and Cellular Biology  
Osaka University  
1-3 Yamada-oka, Suita, Osaka 565, Japan.  
Location/Qualifiers  
1..397  
/organism="Homo sapiens"  
/db\_xref="GDB:D057613E"  
/db\_xref="taxon:9606"  
/map="17"  
/clone="sl73"  
/lab\_host="E.coli"  
/note="3'-directed regional cDNA library. Cleaved by MboI  
and transformed into E.coli."

E COUNT 114 a 80 c 109 g 93 t 1 others  
GIN

Query Match 12.7%; Score 7; DB 20; Length 397;  
Best Local Similarity 100.0%; Pred. No. 5.4e+04;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTCGACG 55  
Db 187 CTCGACG 181

RESULT 11  
LOCUS D12355/c 356 bp mRNA EST 02-DEC-1992  
DEFINITION HUM0007B04 Liver HepG2 cell line. Homo sapiens cDNA clone tb04,  
RNA sequence.

ACCESSION D12355  
NID 62148499  
VERSION D12355.1 GI:2148499  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y.  
and Matsubara, K.  
TITLE Large scale cDNA sequencing for analysis of quantitative and  
qualitative aspects of gene expression  
NATURE GENET. 2, 173-179 (1992)  
94258199

AUTHORS Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y.  
and Matsubara, K.  
TITLE Large scale cDNA sequencing for analysis of quantitative and  
qualitative aspects of gene expression  
JOURNAL MEDLINE 94258199  
COMMENT 2 (bases 1 to 356)  
Murakawa, K., Matsubara, K., Fukushima, A., Yoshii, J. and Okubo, K.  
Chromosomal assignments of 3'-directed partial cDNA sequences  
representing novel genes expressed in granulocytoid cells  
GENOMICS 23, 379-389 (1994)  
95137584

FEATURES  
SOURCE Contact: Kouzaku Okubo, Naohiro Hori, Ryo Matoba, Toshiyuki  
Niiyama, Atsushi Fukushima, Yoko Kojima & Kenichi Matsubara  
Institute for Molecular and Cellular Biology  
Osaka University  
1-3 Yamada-oka, Suita, Osaka 565, Japan.  
Location/Qualifiers  
1..356  
/organism="Homo sapiens"  
/db\_xref="GDB:D057837E"  
/db\_xref="taxon:9606"  
/clone="tb04"  
/lab\_host="E.coli"  
/note="3'-directed regional cDNA library. Cleaved by MboI  
and transformed into E.coli."

BASE COUNT 98 a 99 c 100 g 56 t 3 others  
ORIGIN

Query Match 12.7%; Score 7; DB 20; Length 356;  
Best Local Similarity 100.0%; Pred. No. 5.3e+04;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTCGACG 55  
Db 51 CTCGACG 45

RESULT 12  
LOCUS D15140/c 322 bp mRNA EST 20-JUL-1998  
DEFINITION R1000143A Rice callus Oryza sativa cDNA clone R, RNA sequence.

ACCESSION D15140  
NID 9286332  
VERSION D15140.1 GI:286332  
KEYWORDS EST.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa

REFERENCE 1 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
Poaceae; Oryza.  
AUTHORS 1 (bases 1 to 322)  
Sasaki, T. and Minobe, Y.  
TITLE Rice cDNA from callus  
JOURNAL Rice cDNA from callus  
COMMENT Unpublished (1994)

CONTACT: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program  
2-1-2 Kannondai, Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441  
Fax: 0298-38-7468  
Email: tsasaki@agr.affrc.go.jp  
PROJECT "RGP"  
Seq primer: oligo(dT).  
Location/Qualifiers  
1..322  
/organism="Oryza sativa"  
/strain="cultivar Nipponbare, sub-species Japonica"  
/db\_xref="taxon:4550"  
/clone="R"  
/note="Vector: pBluescript II SK+; Site 1: SalI; S'  
NotI; cDNA prepared from rice callus RNAs by us"

oligo(dT) as a primer and ligating to the SalI-NotI site  
of pBluescript II SK+ phagemid. 10 others

BASE COUNT 53 a 83 c 85 g 91 t  
ORIGIN

Query Match 12.7%; Score 7; DB 20; Length 322;  
Best Local Similarity 100.0%; Pred. No. 5.2e+04;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 49 CTGCACG 55  
|||||||  
Db 275 CTGCACG 269

RESULT 13  
D15143/c 332 bp mRNA EST 20-JUL-1998  
LOCUS R1CC0149A Rice callus Oryza sativa cDNA clone R, mRNA sequence.  
DEFINITION R1CC0149A Rice callus Oryza sativa cDNA clone R, mRNA sequence.  
SESSION D15143  
KEYWORDS 9286335  
D15143.1 GI:286335  
EST.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa.  
REFERENCE 1 (bases 1 to 332)  
AUTHORS Sasaki,T. and Minobe,Y.  
TITLE Rice cDNA from callus  
JOURNAL Unpublished (1994)  
COMMENT

Contact: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program  
2-1-2 Kannondai, Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441  
Fax: 0298-38-7468  
Email: tsasaki@abr.affrc.go.jp  
PROJECT = RGP  
Seq primer: oligo(dT).  
Location/Qualifiers  
1. .332  
/organism="Oryza sativa"  
/strain="cultivar Nipponbare, sub-species Japonica"  
/db\_xref="taxon:4530"  
/clone="R"  
/clone\_id="Rice callus"  
/note="Vector: pBluescript II SK+; Site\_1: SalI; Site\_2:  
NotI; cDNA prepared from rice callus mRNAs by using  
oligo(dT) as a primer and ligating to the SalI-NotI site  
of pBluescript II SK+ phagemid. 2 others

BASE COUNT 69 a 85 c 96 g 80 t  
ORIGIN

Query Match 12.7%; Score 7; DB 20; Length 332;  
Best Local Similarity 100.0%; Pred. No. 5.3e+04;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 49 CTGCACG 55  
|||||||  
Db 76 CTGCACG 70

RESULT 14  
D15154 273 bp mRNA EST 20-JUL-1998  
LOCUS R1CC0177A Rice callus Oryza sativa cDNA clone R, mRNA sequence.  
DEFINITION R1CC0177A Rice callus Oryza sativa cDNA clone R, mRNA sequence.  
ACCESSION D15154

NID 9286345  
VERSION D15154.1 GI:286345  
KEYWORDS EST.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa.

REFERENCE 1 (bases 1 to 273)  
AUTHORS Sasaki,T. and Minobe,Y.  
TITLE Rice cDNA from callus  
JOURNAL Unpublished (1994)  
COMMENT

Contact: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program  
2-1-2 Kannondai, Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441  
Fax: 0298-38-7468  
Email: tsasaki@abr.affrc.go.jp  
PROJECT = RGP  
Seq primer: oligo(dT).  
Location/Qualifiers  
1. .273  
/organism="Oryza sativa"  
/strain="cultivar Nipponbare, sub-species Japonica"  
/db\_xref="taxon:4530"  
/clone="R"  
/clone\_id="Rice callus"  
/note="Vector: pBluescript II SK+; Site\_1: SalI; Site\_2:  
NotI; cDNA prepared from rice callus mRNAs by using  
oligo(dT) as a primer and ligating to the SalI-NotI site  
of pBluescript II SK+ phagemid. 11 others

BASE COUNT 43 a 87 c 72 g 60 t  
ORIGIN

Query Match 12.7%; Score 7; DB 20; Length 273;  
Best Local Similarity 100.0%; Pred. No. 5.1e+04;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 49 CTGCACG 55  
|||||||  
Db 85 CTGCACG 91

RESULT 15  
D15173/c 356 bp mRNA EST 20-JUL-1998  
LOCUS R1CC0205A Rice callus Oryza sativa cDNA clone R, mRNA sequence.  
DEFINITION R1CC0205A Rice callus Oryza sativa cDNA clone R, mRNA sequence.  
ACCESSION D15173  
NID 9286365  
VERSION D15173.1 GI:286365  
KEYWORDS EST.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa.

REFERENCE 1 (bases 1 to 356)  
AUTHORS Sasaki,T. and Minobe,Y.  
TITLE Rice cDNA from callus  
JOURNAL Unpublished (1994)  
COMMENT

Contact: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program  
2-1-2 Kannondai, Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441

Fax: 0298-38-7468  
 Email: tsasakieabr.affrc.go.jp  
 PROJECT = 'RGP'  
 Seq primer: oligo(dT).  
 Location/Qualifiers  
 1. 356

FEATURES  
 Source  
 /organism="Oryza sativa"  
 /strain="cultivar Nipponbare, sub\_species Japonica"  
 /db\_xref="taxon:4530"  
 /clone="R"  
 /clone\_id="Rice callus"  
 /note="Vector: pBluescript II SK+; Site\_1: SalI; Site\_2:  
 NotI; cDNA prepared from rice callus mRNAs by using  
 oligo(dT) as a primer and ligating to the SalI-NotI site  
 of pBluescript II SK+ phagemid."  
 BASE COUNT 71 a 88 c 111 g 75 t 11 others  
 ORIGIN

Query Match 12.7%; Score 7; DB 20; Length 356;  
 Best Local Similarity 100.0%; Pred. No. 5.3e+04;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 CTGCACG 55  
 |||||  
 DB 102 CTGCACG 96

Search completed: September 17, 1999, 21:27:52  
 Job time: 14256 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:27:56 ; Search time 2825.05 Seconds  
(without alignments)  
6.982 Million cell updates/sec

Title: US-08-956-518a-87  
Perfect score: 10  
Sequence: 1 TGTGTGTACAG 10

## Scoring table:

IDENTITY\_NUC 2546578 seqs, 98626752 residues

## Database:

EST:  
1: em\_est1:  
2: em\_est2:  
3: em\_est3:  
4: em\_est4:  
5: em\_est5:  
6: em\_est6:  
7: em\_est7:  
8: em\_est8:  
9: em\_est9:  
10: em\_est10:  
11: em\_est11:  
12: em\_est12:  
13: em\_est13:  
14: em\_est14:  
15: em\_est15:  
16: em\_est16:  
17: em\_est17:  
18: em\_est18:  
19: em\_est19:  
20: em\_est20:  
21: em\_est21:  
22: em\_est22:  
23: em\_est23:  
24: em\_est24:  
25: em\_est25:  
26: em\_est26:  
27: em\_est27:  
28: em\_est28:  
29: em\_est29:  
30: em\_est30:  
31: em\_est31:  
32: em\_est32:  
33: em\_est33:  
34: em\_est34:  
35: em\_est35:  
36: em\_est36:  
37: em\_est37:  
38: em\_est38:  
39: em\_est39:  
40: em\_est40:  
41: em\_est41:  
42: em\_est42:  
43: em\_est43:  
44: em\_est44:  
45: em\_est45:  
46: em\_est46:  
47: em\_est47:  
48: em\_est48:  
49: em\_est49:  
50: em\_est50:  
51: em\_est51:  
52: em\_est52:  
53: em\_est53:

54: em\_est22:  
55: em\_est23:  
56: em\_est24:  
57: em\_est25:  
58: em\_est26:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100.0	429	5	AF120344	AF120344 Homo sapi
2	100.0	389	20	D12193	D12193 HUM000317
3	100.0	566	20	D20010	D20010 HUM000981
4	100.0	88	20	D20990	D20990 HUM001972
5	100.0	199	20	D31098	D31098 HUM012503 H
6	100.0	374	20	D32407	D32407 CELK0136R
7	100.0	405	20	D33452	D33452 CELK031BR
8	100.0	344	20	F02697	F02697 HSC17B112 n
9	100.0	265	20	F03251	F03251 HSC1RA102 n
10	100.0	314	20	M78805	M78805 EST00953 H1
11	100.0	290	20	M78839	M78839 EST00987 H1
12	100.0	332	20	M78856	M78856 EST01004 H1
13	100.0	390	20	M79816	M79816 WES000353 M
14	100.0	225	20	M85650	M85650 EST02167 Fe
15	100.0	345	20	M86021	M86021 EST02546 Fe
16	100.0	247	20	M86024	M86024 EST02549 Fe
17	100.0	526	20	M88899	M88899 CEL10F4 Chr
18	100.0	312	20	T02241	T02241 WES002962 E
19	100.0	315	20	T02245	T02245 WES002966 E
20	100.0	408	20	T07350	T07350 EST05239 Fe
21	100.0	370	20	T08327	T08327 EST06218 In
22	100.0	429	20	T08749	T08749 EST06641 In
23	100.0	388	20	T09101	T09101 EST06994 In
24	100.0	393	20	T11545	T11545 A1418F Hear
25	100.0	214	20	T14734	T14734 05-04-f11-f2
26	100.0	248	20	T15405	T15405 IB1093 Infa
27	100.0	343	20	T15545	T15545 IB1523 Infa
28	100.0	415	20	T16583	T16583 NIB1514 NOR
29	100.0	415	20	T17119	T17119 NIB352 Norm
30	100.0	337	20	T20645	T20645 2653 Lambda
31	100.0	518	20	T26572	T26572 AB324D9F Id
32	100.0	450	20	T27084	T27084 NIB1261G03R
33	100.0	342	20	T28031	T28031 EST24388 Hu
34	100.0	346	20	T28461	T28461 EST45746 Hu
35	100.0	321	20	T29601	T29601 EST86422 Hu
36	100.0	326	20	T30869	T30869 EST24325 Hu
37	100.0	336	20	T33366	T33366 EST47555 Hu
38	100.0	336	20	T33367	T33367 EST47556 Hu
39	100.0	333	20	T33499	T33499 EST58071 Hu
40	100.0	488	20	T33648	T33648 EST58566 Hu
41	100.0	360	20	T34126	T34126 EST62996 Hu
42	100.0	221	20	T34206	T34206 EST64188 Hu
43	100.0	258	20	T35232	T35232 EST81801 Hu
44	100.0	263	20	T35321	T35321 EST82821 Hu
45	100.0	495	54	HSW009549	A1044699 Homo sapi

## ALIGNMENTS

RESULT 1  
ID AF120344  
XX AF120344 standard; RNA; EST; 429 BP.  
AC AF120344;  
XX  
SV AF120344.1  
XX



NI 94325197  
 XX 11-MAR-1999 (Rel. 59, Created)  
 DT 11-MAR-1999 (Rel. 59, Last updated, Version 1)  
 XX  
 DE Homo sapiens clone IEDP1-29, mRNA sequence.  
 XX EST.  
 KW  
 XX Homo sapiens (human)  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;  
 OC Primates; Catarrhini; Homnidae; Homo.  
 XX  
 XX [1]  
 RP 1-429  
 RA MEDLINE; 97471696.  
 RA Jacob A.N., Baskaran N., Kandpal G., Narayan D., Bhargava A.K.,  
 Kandpal R.P.;  
 "Isolation of human ear specific cDNAs and construction of cDNA  
 libraries from surgically removed small amounts of inner ear tissues";  
 Somat. Cell Mol. Genet. 23(2):83-95(1997).  
 XX  
 XX [2]  
 RP 1-429  
 RA Nellisery J.K., Kandpal R.P.;  
 Submitted (15-JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RL Fels Institute for Cancer Research, Temple University School of  
 Medicine, 3307 North Broad Street, Philadelphia, PA 19140, USA  
 XX  
 XX Key Location/Qualifiers  
 FH 1.429  
 FT /db\_xref="taxon:9606"  
 FT /note="Isolated by kinetic enrichment in a representational  
 difference analysis (RDA) experiment."  
 FT /organism="Homo sapiens"  
 FT /tissue-type="inner and middle ear"  
 FT /dev-stage="fetus"  
 FT /clone="IEDP1-29"  
 XX  
 SO Sequence 429 bp; 146 A; 58 C; 85 G; 134 T; 6 other;

Query Match 100.0%; Score 10; DB 5; Length 429;  
 Best Local Similarity 100.0%; Pred. No. 5.5e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 TGTGTCTCAG 10  
 244 TGTGTCTCAG 253

RESULT 2  
 LOCUS D12193 389 bp mRNA EST 07-OCT-1996  
 DEFINITION HUM0005317 Liver HepG2 cell line. Homo sapiens CDNA clone s317 3',  
 mRNA sequence.  
 ACCESSION D12193  
 MID 9767720  
 VERSION D12193.1 GI:767720  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 389)  
 Okubo,K., Hori,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y.  
 and Matsubara,K.  
 Large scale cDNA sequencing for analysis of quantitative and  
 qualitative aspects of gene expression  
 JOURNAL Nature Genet. 2, 173-179 (1992)  
 MEDLINE 94258199

REFERENCE 2 (bases 1 to 389)  
 AUTHORS Murakawa,K., Matsubara,K., Fukushima,A., Yoshii,J. and Okubo,K.  
 TITLE Chromosomal assignments of 3'-directed partial cDNA sequences  
 representing novel genes expressed in granulocytoid cells  
 JOURNAL Genomics 23, 379-389 (1994)  
 MEDLINE 95137584  
 COMMENT Contact: Kousaku Okubo, Naohiro Hori, Ryo Matoba, Toshiyuki  
 Niiyama, Atsushi Fukushima, Yoko Kojima & Kenichi Matsubara  
 Institute for Molecular and Cellular Biology  
 Osaka University  
 1-3 Yamada-oka, Suita, Osaka 565, Japan.  
 FEATURES  
 source Location/Qualifiers  
 1.389  
 /organism="Homo sapiens"  
 /db\_xref="CDB:D087694E"  
 /db\_xref="taxon:9606"  
 /map="14"  
 /clone="s317"  
 /clone\_1lb="Liver HepG2 cell line."  
 /lab\_host="E.coli"  
 /note="3'-directed regional cDNA library. Cleaved by MboI  
 and transformed into E.coli."  
 BASE COUNT 101 a 94 c 70 g 124 t  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 389;  
 Best Local Similarity 100.0%; Pred. No. 5.5e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 TGTGTCTCAG 10  
 286 TGTGTCTCAG 295

RESULT 3  
 LOCUS D20010 566 bp mRNA EST 30-JUL-1996  
 DEFINITION HUMGS00981 Human promyelocyte Homo sapiens CDNA clone pm2779 3',  
 mRNA sequence.  
 ACCESSION D20010  
 MID 9500907  
 VERSION D20010.1 GI:500907  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 566)  
 Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,  
 Yoshinari,H., Arimoto,J. and Matsubara,K.  
 Gene expression of human promyelocytic cell line HL60 before and  
 after induction of differentiation. A new application of 3'directed  
 cDNA sequencing  
 Unpublished (1993)  
 JOURNAL  
 COMMENT Contact: Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,  
 Yoshinari,H., Arimoto,J. and Matsubara,K.  
 Institute for Molecular and Cellular Biology  
 Osaka University  
 3-1 Yamada-oka, Suita, Osaka 565, Japan.  
 FEATURES  
 source Location/Qualifiers  
 1.566  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="pm2779"  
 /clone\_1lb="Human promyelocyte"  
 /note="Female, adult, cell\_line = HL60, cell\_type =  
 promyelocyte."  
 BASE COUNT 138 a 114 c 118 g 176 t 20 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 566;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGTGTCTAG 10  
 |||||  
 Db 164 TGTGTGTCTAG 173

RESULT 4  
 D20990/c 88 bp mRNA EST 30-JUL-1996  
 LOCUS HUMGS01972 Human promyelocyte Homo sapiens cDNA clone mp2519.3',  
 DEFINITION mRNA sequence.  
 ACCESSION D20990  
 NID 9504810  
 VERSION D20990.1 GI:504810  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Euteria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 88)  
 Okubo, K., Fukushima, A., Yoshii, J., Niiyama, T., Kojima, Y.,  
 Yoshinari, H., Arimoto, J. and Matsubara, K.,  
 Gene expression of human promyelocytic cell line HL60 before and  
 after induction of differentiation. A new application of 3'directed  
 cDNA sequencing  
 JOURNAL Unpublished (1993)  
 COMMENT

FEATURES  
 SOURCE  
 1. 88  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="mp2519"  
 /note="Female, adult, cell\_line = HL60, cell\_type =  
 promyelocyte." 12 g 32 t 1 others

BASE COUNT 23 a 20 c 12 g 32 t 1 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 88;  
 Best Local Similarity 100.0%; Pred. No. 5.3e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGTGTCTAG 10  
 |||||  
 Db 52 TGTGTGTCTAG 43

RESULT 5  
 D31098 199 bp mRNA EST 08-FEB-1995  
 LOCUS HUMH12503 Human fetal lung Homo sapiens cDNA 5', mRNA sequence.  
 DEFINITION  
 ACCESSION D31098  
 NID 9643978  
 VERSION D31098.1 GI:643978  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Euteria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 199)  
 Sudu, K., Chinen, K. and Nakamura, Y.,  
 2058 expressed sequence tags (ESTs) from a human fetal lung cDNA  
 library

JOURNAL Genomics 24, 276-279 (1995)  
 MEDLINE 95213017  
 COMMENT  
 Contact: Yusuke Nakamura  
 Institute of Medical Science  
 University of Tokyo  
 4-6-1 Shirokanedai, Minato-ku, Tokyo 108, Japan  
 Tel: 81-3-5449-5372  
 Fax: 81-3-5449-5433  
 Email: yusuke@ims.u-tokyo.ac.jp  
 Insert Length: 552 Std Error: 0.00  
 High quality sequence stop: 1.  
 Location/Qualifiers  
 1. 199  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="17q21"  
 /clone\_lib="Human fetal lung"

BASE COUNT 58 a 40 c 36 g 59 t 6 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 199;  
 Best Local Similarity 100.0%; Pred. No. 5.4e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGTGTCTAG 10  
 |||||  
 Db 162 TGTGTGTCTAG 171

RESULT 6  
 D32407 374 bp mRNA EST 08-AUG-1994  
 LOCUS CELK01366R Yuj1 Kohara unpublished cDNA Caenorhabditis elegans cDNA  
 DEFINITION clone yk1366.3', mRNA sequence.  
 ACCESSION D32407  
 NID 9523368  
 VERSION D32407.1 GI:523368  
 KEYWORDS EST.  
 SOURCE  
 ORGANISM Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditida;  
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 REFERENCE 1 (bases 1 to 374)  
 Kohara, Y., Mitsuki, H., Nishigaki, A., Motonashi, T., Sugimoto, A. and  
 Tabara, H.,  
 Toward an expression map of the C. elegans genome  
 JOURNAL Unpublished (1994)  
 COMMENT

FEATURES  
 SOURCE  
 1. 374  
 /organism="Caenorhabditis elegans"  
 /strain="CB1489 him-8(e1489)"  
 /db\_xref="taxon:6239"  
 /map="978E04; 1"  
 /clone="YK1366"  
 /clone\_lib="Yuj1 Kohara unpublished cDNA"  
 /note="dev\_stage=varied, sex=Hermaphrodite male,  
 tissue\_type=whole animal" 70 c 67 g 118 t

BASE COUNT 119 a 70 c 67 g 118 t  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 5.5e+03;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TGTGTGTCAG 10  
 |||||||||  
 Db 196 TGTGTGTCAG 205

## RESULT 7

D33452 406 bp mRNA EST 08-AUG-1994  
 LOCUS CEK031B8 Yuj1 Kohara unpublished cDNA Caenorhabditis elegans cDNA  
 DEFINITION clone YK31B8 3', mRNA sequence.

ACCESSION D33452  
 NID 9524372  
 VERSION D33452.1 GI:524372

KEYWORDS EST.  
 SOURCE Caenorhabditis elegans.

## ORGANISM

Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderiinae; Caenorhabditis.  
 1 (bases 1 to 406)

## REFERENCE

Kohara,Y., Mitsukl,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and  
 Tabara,H.

Toward an expression map of the C.elegans genome

## JOURNAL

Unpublished (1994)  
 On Sep 21, 1992 this sequence version replaced gi:279017.

## COMMENT

Contact: Yuj1 Kohara  
 Gene Library Lab  
 National Institute of Genetics  
 Yata 1111, Mishima, Shizuoka 411, Japan  
 Tel: 0559-75-0771  
 Fax: 0559-75-6240  
 Email: ykoha@eddb.nig.ac.jp.

## FEATURES

## SOURCE

Location/Qualifiers  
 1..406

/organism="Caenorhabditis elegans"

/strain="CB1489 him-8(e1489)"

/db\_xref="taxon:6239"

/clone="YK31B8"

/clone\_lib="Yuj1 Kohara unpublished cDNA"

/note="dev-stage=varied, sex=Hermaphrodite male,  
 tissue=whole animal"

## BASE COUNT

162 a 40 c 100 g 104 t

## ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 406;  
 Best Local Similarity 100.0%; Pred. No. 5.5e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTCAG 10  
 |||||||||

Db 328 TGTGTGTCAG 337

## RESULT 8

F02697 344 bp mRNA EST 02-FEB-1995  
 LOCUS HSC17B12 normalized infant brain cDNA Homo sapiens cDNA clone  
 DEFINITION c-17b11 3', mRNA sequence.

## ACCESSION

## NID

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Homo sapiens  
 human.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

## REFERENCE

Aufiray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,  
 Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,  
 Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,

Sebastiani-Kabakchis,C. and Tessier,A.  
 IMAGE: molecular integration of the analysis of the human genome  
 and its expression  
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
 95277534

## COMMENT

Contact: Genethon  
 Genexpress-Genethon  
 Genethon Centre de recherche sur le Genome Humain  
 1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE  
 Tel: 33169472800  
 Fax: 33160778698

Email: genexpress@genethon.fr  
 Single read, removed at sequence 5'end  
 Genexpress\_library\_id: C; Genexpress\_sequence\_id: alc-17b11  
 Seq primer: (-21)M13-universal..

## FEATURES

## SOURCE

## ORIGIN

## BASE COUNT

## ORIGIN

115 a 71 c 55 g 101 t 2 others  
 Query Match 100.0%; Score 10; DB 20; Length 344;  
 Best Local Similarity 100.0%; Pred. No. 5.5e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTCAG 10  
 |||||||||

Db 45 TGTGTGTCAG 36

## RESULT 9

F03251 265 bp mRNA EST 02-FEB-1995  
 LOCUS HSC1A102 normalized infant brain cDNA Homo sapiens cDNA clone  
 DEFINITION c-1ra10 3', mRNA sequence.

## ACCESSION

## NID

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## COMMENT

Contact: Genethon  
 Genexpress-Genethon

## ORIGIN

## BASE COUNT

## ORIGIN

## BASE COUNT

## ORIGIN

## BASE COUNT

## ORIGIN

## BASE COUNT

## ORIGIN

## BASE COUNT

## ORIGIN

## BASE COUNT

## ORIGIN

## BASE COUNT

## ORIGIN

## BASE COUNT

## ORIGIN

## BASE COUNT

## ORIGIN

## BASE COUNT

## ORIGIN

## BASE COUNT

## ORIGIN

Genethon Centre de recherche sur le Genome Humain  
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE  
Tel: 33169472800  
Fax: 33160778698  
Email: genexpress@genethon.fr  
Single read. removed at sequence 5' end  
Genexpress library idt: C; Genexpress\_sequence\_idt: alc-lra10  
Seq primer: (-21)M13-universal.

## FEATURES

## source

Location/Qualifiers  
1. 265  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="17q21"  
/clone="c-lra10"  
/clone\_lib="normalized infant brain cDNA"  
/sex="Female"  
/tissue\_type="total brain"  
/dev\_stage="3 months old"  
/note="Organ: brain; Vector: lambda BA; Site: 1: HindIII;  
Site 2: NotI; sex=Female; dev\_stage=3 months old;  
Isolate-muscular atrophy patient; tissue\_type=total  
brain; total mRNA was oligo-(dT) primed and directionally  
cloned 5' -> 3' into the HindIII -> NotI sites of the  
lambda BA vector. Clone library from B. Soares, Psychiatry  
Dept. Columbia University, USA. Normalization method:  
Bento Soares, P.N.A.S. in press"

## BASE COUNT

78 a 33 c 71 g 75 t 8 others

## ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 265;  
Best Local Similarity 100.0%; Pred. No. 5.5e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTGAC 10  
|||||

Db 227 TGTGTGTGAC 236

## RESULT 10

M78805 314 bp mRNA EST 26-MAY-1992  
LOCUS EST00953 Hippocampus, Stratagene (cat. #936205) Homo sapiens cDNA  
DEFINITION clone HHC602, mRNA sequence.

ACCESSION M78805  
NID 9273120  
VERSION M78805.1 GI:273120  
KEYWORDS EST.  
RCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 314)

AUTHORS Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,  
Utterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.  
TITLE Sequence identification of 2,375 human brain genes  
JOURNAL Nature 355, 632-634 (1992)  
MEDLINE 92168112

REFERENCE 2 (bases 1 to 314)  
AUTHORS Polymeropoulos,M.H., Xiao,H., Sikela,J.M., Adams,M., Venter,J.C.  
and Merrill,C.R.  
TITLE Chromosomal distribution of 320 genes from a brain cDNA library  
JOURNAL Nature Genet. 4, 381-386 (1993)  
MEDLINE 94004966

## COMMENT

Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
Seq primer: M13 Forward.

## FEATURES

## source

Location/Qualifiers  
1. 314  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):78249"  
/db\_xref="GDB:D12S110E"  
/db\_xref="taxon:9606"  
/map="12"  
/clone="HHC602"  
/clone\_lib="Hippocampus, Stratagene (cat. #936205)"  
/note="Vector: lambdaZAP-II; Female, 2 years; oligo-dT +  
random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb  
average insert size."

## BASE COUNT

70 a 67 c 80 g 94 t 3 others

## ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 314;  
Best Local Similarity 100.0%; Pred. No. 5.5e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTGAC 10  
|||||

Db 35 TGTGTGTGAC 26

## RESULT 11

M78839 290 bp mRNA EST 26-MAY-1992  
LOCUS EST00987 Hippocampus, Stratagene (cat. #936205) Homo sapiens cDNA  
DEFINITION clone HHC695, mRNA sequence.

ACCESSION M78839  
NID 9273154  
VERSION M78839.1 GI:273154  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 290)

AUTHORS Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,  
Utterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.  
TITLE Sequence identification of 2,375 human brain genes  
JOURNAL Nature 355, 632-634 (1992)  
MEDLINE 92168112

REFERENCE 2 (bases 1 to 290)  
AUTHORS Polymeropoulos,M.H., Xiao,H., Sikela,J.M., Adams,M., Venter,J.C.  
and Merrill,C.R.  
TITLE Chromosomal distribution of 320 genes from a brain cDNA library  
JOURNAL Nature Genet. 4, 381-386 (1993)  
MEDLINE 94004966

## COMMENT

Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
Seq primer: M13 Forward.

## FEATURES

## source

Location/Qualifiers  
1. 290  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):78292"  
/db\_xref="GDB:D10S238E"  
/db\_xref="taxon:9606"  
/map="10"  
/clone="HHC695"

/clone\_lib="Hippocampus, Stratagene (cat. #936205)"  
/note="Vector: lambdaZAP-II; Female, 2 years; oligo-dT +  
random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb  
average insert size."

## BASE COUNT

76 a 99 c 52 g 61 t 2 others

## ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 290;  
 Best Local Similarity 100.0%; Pred. No. 5.5e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGTCTCAG 10  
 |||||  
 Db 245 TGTGTCTCAG 236

RESULT 12  
 M78856 332 bp mRNA EST 26-MAY-1992  
 LOCUS EST01004 Hippocampus, Striatogene (cat. #936205) Homo sapiens CDNA  
 DEFINITION clone HRCMH72, mRNA sequence.  
 ACCESSION M78856  
 9273171  
 .ION M78856.1 GI:273171  
 .MORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 332)  
 AUTHORS Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,  
 Uterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.  
 TITLE Sequence identification of 2,375 human brain genes  
 JOURNAL Nature 355, 632-634 (1992)  
 MEDLINE 92168112  
 COMMENT Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org  
 Seq primer: M13 forward.  
 Location/Qualifiers  
 1. .332  
 /organism="Homo sapiens"  
 /db\_xref="Arcc (Inhost):78321"  
 /db\_xref="GDB:DS1837E"  
 /db\_xref="taxon:9606"  
 /clone="HRCMH72"  
 /note="Vector: lambdaZAP-II; Female, 2 Years; oligo-dt +  
 random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb  
 average insert size."  
 BASE COUNT 91 a 111 c 59 g 71 t  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 332;  
 Best Local Similarity 100.0%; Pred. No. 5.5e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGTCTCAG 10  
 |||||  
 Db 107 TGTGTCTCAG 98

RESULT 13  
 M79816/c 390 bp mRNA EST 30-JUN-1992  
 LOCUS EST00353 Mixed stage, Striatogene (cat. #937006) Caenorhabditis  
 DEFINITION elegans CDNA cEMSE23, mRNA sequence.  
 ACCESSION M79816  
 9271835  
 .ION M79816.1 GI:271835  
 .MORDS EST.  
 SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;  
 Rhabditina; Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.  
 REFERENCE 1 (bases 1 to 390)  
 AUTHORS McCombie,W.R., Adams,M.D., Kelley,J.M., Fitzgerald,M.G.,  
 Uterback,T.R., Khan,M., Dubnick,M., Kerlavage,A.R., Venter,J. and  
 Fields,C.  
 TITLE Caenorhabditis elegans expressed sequence tags identify gene  
 families and potential disease gene homologues  
 JOURNAL Nature Genet. 1, 124-131 (1992)  
 MEDLINE 93250983  
 COMMENT Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org  
 Seq primer: M13 forward.  
 Location/Qualifiers  
 1. .390  
 /organism="Caenorhabditis elegans"  
 /db\_xref="taxon:6239"  
 /clone="CENSE23"  
 /note="Vector: Uni-ZAP XR; C. elegans mixed stage CDNA  
 library. Striatogene catalog #937006. The library is oligo  
 dt primed and directionally cloned in the Uni-ZAP XR  
 vector."  
 BASE COUNT 127 a 67 c 70 g 120 t 6 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 390;  
 Best Local Similarity 100.0%; Pred. No. 5.5e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGTCTCAG 10  
 |||||  
 Db 139 TGTGTCTCAG 130

RESULT 14  
 M85650 225 bp mRNA EST 26-MAY-1992  
 LOCUS EST02167 Fetal brain, Striatogene (cat.#936206) Homo sapiens CDNA  
 DEFINITION clone HFC138, mRNA sequence.  
 ACCESSION M85650  
 9274297  
 .ION M85650.1 GI:274297  
 .MORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 225)  
 AUTHORS Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,  
 Uterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.  
 TITLE Sequence identification of 2,375 human brain genes  
 JOURNAL Nature 355, 632-634 (1992)  
 MEDLINE 92168112  
 COMMENT Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org  
 Seq primer: M13 forward.  
 Location/Qualifiers  
 1. .225  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="HFC138"  
 /note="Vector: lambdaZAP-II; Male, 2 Years; oligo-dt +  
 random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb  
 average insert size."  
 BASE COUNT 127 a 67 c 70 g 120 t 6 others  
 ORIGIN

Search completed: September 17, 1999, 21:27:57  
 Job time: 14261 sec

```

/organism="Homo sapiens"
/db_xref="ATCC (lnhost):81227"
/db_xref="GDB:D0S343E"
/db_xref="taxon:9606"
/clone="HFBCL38"
/clone_1lb="Fetal brain, Stratagene (cat#936206)"
/note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
oligo-dt + random primed cDNA synthesis; lambdaZAP-II
vector, 1.0kb average insert size."
BASE COUNT      67 a      67 c      49 g      41 t      1 others
ORIGIN

```

```

Query Match      100.0%; Score 10; DB 20; Length 225;
Best Local Similarity 100.0%; Pred. No. 5.4e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 TGTGTCTCAG 10
        |||||
        66 TGTGTCTCAG 57

```

```

RESULT 15
M86021      345 bp      mRNA      EST      26-MAY-1992
LOCUS      EST02546 Fetal brain, Stratagene (cat#936206) Homo sapiens CDNA
DEFINITION clone HFB0Y34, mRNA sequence.
ACCESSION  M86021
NID        9274672
VERSION    M86021.1 GI:274672
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 345)
AUTHORS   Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,
            Utterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.
            Sequence identification of 2,375 human brain genes
            Nature 355, 632-634 (1992)
            92168112
JOURNAL
MEDLINE
COMMENT    Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: atkerlav@tigr.org
            Seq primer: M13 Forward.
            Location/Qualifiers

```

```

FEATURES
    source
        1..345
            /organism="Homo sapiens"
            /db_xref="ATCC (lnhost):81598"
            /db_xref="GDB:D0S709E"
            /db_xref="taxon:9606"
            /clone="HFB0Y34"
            /clone_1lb="Fetal brain, Stratagene (cat#936206)"
            /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
            oligo-dt + random primed cDNA synthesis; lambdaZAP-II
            vector, 1.0kb average insert size."
BASE COUNT      110 a      68 c      63 g      104 t
ORIGIN

```

```

Query Match      100.0%; Score 10; DB 20; Length 345;
Best Local Similarity 100.0%; Pred. No. 5.5e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 TGTGTCTCAG 10
        |||||
        DB      18 TGTGTCTCAG 27

```

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:27:55 : Search time 2825.05 Seconds  
(without alignments)  
6.982 Million cell updates/sec

Title: US-08-956-518a-86  
Perfect score: 10  
Sequence: 1 TTTTGAAG 10

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 98626752 residues

Database:

EST:  
1: em\_est1.\*  
2: em\_est2.\*  
3: em\_est3.\*  
4: em\_est4.\*  
5: em\_est5.\*  
6: em\_est6.\*  
7: em\_est7.\*  
8: em\_est8.\*  
9: em\_est9.\*  
10: em\_est10.\*  
11: em\_est11.\*  
12: em\_est12.\*  
13: em\_est13.\*  
14: em\_est14.\*  
15: em\_est15.\*  
16: em\_est16.\*  
17: em\_est17.\*  
18: em\_est18.\*  
19: em\_est19.\*  
20: em\_est20.\*  
21: em\_est21.\*  
22: em\_est22.\*  
23: em\_est23.\*  
24: em\_est24.\*  
25: em\_est25.\*  
26: em\_est26.\*  
27: em\_est27.\*  
28: em\_est28.\*  
29: em\_est29.\*  
30: em\_est30.\*  
31: em\_est31.\*  
32: em\_est32.\*  
33: em\_est33.\*  
34: em\_est34.\*  
35: em\_est35.\*  
36: em\_est36.\*  
37: em\_est37.\*  
38: em\_est38.\*  
39: em\_est39.\*  
40: em\_est40.\*  
41: em\_est41.\*  
42: em\_est42.\*  
43: em\_est43.\*  
44: em\_est44.\*  
45: em\_est45.\*  
46: em\_est46.\*  
47: em\_est47.\*  
48: em\_est48.\*  
49: em\_est49.\*  
50: em\_est50.\*  
51: em\_est51.\*  
52: em\_est52.\*  
53: em\_est53.\*

54: em\_est22.\*  
55: em\_est23.\*  
56: em\_est24.\*  
57: em\_est25.\*  
58: em\_est26.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	10	100.0	540	5 AF123245	AF123245 Gossypium
C 2	10	100.0	314	20 D20170	D20170 HUNG501144
C 3	10	100.0	288	20 D23028	D23028 RICC10206A
C 4	10	100.0	328	20 D22533	D22533 RICC01908 R
C 5	10	100.0	261	20 D23618	D23618 RICC3103A R
C 6	10	100.0	331	20 D24868	D24868 RICR2677A R
C 7	10	100.0	423	20 D26922	D26922 CELK014D2R
C 8	10	100.0	398	20 D27177	D27177 CELK004D5R
C 9	10	100.0	471	20 D27179	D27179 CELK002GXR
C 10	10	100.0	461	20 D27520	D27520 CELK012B4R
C 11	10	100.0	417	20 D27544	D27544 CELK013A7R
C 12	10	100.0	403	20 D27856	D27856 CELK005B2R
C 13	10	100.0	387	20 D27881	D27881 CELK005GXR
C 14	10	100.0	470	20 D27882	D27882 CELK005GXR
C 15	10	100.0	403	20 D27887	D27887 CELK005G2R
C 16	10	100.0	365	20 D28006	D28006 CELK007F6R
C 17	10	100.0	445	20 D28019	D28019 CELK008A1R
C 18	10	100.0	300	20 D31015	D31015 HUN112171 H
C 19	10	100.0	382	20 D32317	D32317 CELK010E5R
C 20	10	100.0	352	20 D32380	D32380 CELK012D2R
C 21	10	100.0	417	20 D32546	D32546 CELK018E8R
C 22	10	100.0	397	20 D32581	D32581 CELK019A6R
C 23	10	100.0	323	20 D32642	D32642 CELK019H3R
C 24	10	100.0	358	20 D32679	D32679 CELK001F4R
C 25	10	100.0	269	20 D32686	D32686 CELK020A5R
C 26	10	100.0	374	20 D32706	D32706 CELK020C9R
C 27	10	100.0	402	20 D32964	D32964 CELK024B7R
C 28	10	100.0	382	20 D32972	D32972 CELK024C5R
C 29	10	100.0	295	20 D33053	D33053 CELK025E2R
C 30	10	100.0	380	20 D33071	D33071 CELK025G7R
C 31	10	100.0	382	20 D33081	D33081 CELK026A9R
C 32	10	100.0	373	20 D33087	D33087 CELK026A9R
C 33	10	100.0	395	20 D33182	D33182 CELK027D9R
C 34	10	100.0	355	20 D33437	D33437 CELK031A9R
C 35	10	100.0	373	20 D33477	D33477 CELK031FXR
C 36	10	100.0	410	20 D33502	D33502 CELK031H7R
C 37	10	100.0	416	20 D33820	D33820 CELK036F3R
C 38	10	100.0	378	20 D33943	D33943 CELK038E9R
C 39	10	100.0	412	20 D33948	D33948 CELK038F4R
C 40	10	100.0	388	20 D34048	D34048 CELK040A8R
C 41	10	100.0	387	20 D34110	D34110 CELK041B4R
C 42	10	100.0	406	20 D34303	D34303 CELK044A3R
C 43	10	100.0	224	20 D34370	D34370 CELK045A2R
C 44	10	100.0	380	20 D34392	D34392 CELK045C6R
C 45	10	100.0	448	54 HSM009362	A1044512 Homo sapi

## ALIGNMENTS

RESULT 1  
ID AF123245/c  
XX AF123245 standard; RNA; EST; 540 BP.  
AC AF123245:  
XX  
SV AF123245.1  
XX

**Nl**      94325284  
**XX**

**DT**    11-MAR-1999 (Rel. 59, Created)  
**XT**    11-MAR-1999 (Rel. 59, Last updated; Version 1)

**DE**    Gossypium hirsutum clone CfI6S strain Texas Marker-1, mRNA sequence.

**KW**    EST.  
**KX**    .  
**XX**    .

**OS**    Gossypium hirsutum (upland cotton)  
**OC**    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
     eumhyllophytes; Spermatophyta; Magnoliophyta; euclcotyledons;  
     core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.

**XX**    [1]  
**RN**    1-540  
**RP**    Cui X., Brown R.M.;  
     "CDNA sequences from secondary-stage cotton fibers";  
     Unpublished.

**XX**    [2]  
**RN**    1-540  
**RP**    Cui X., Brown R.M.;

**RL**    Submitted (23-JAN-1999) to the EMBL/Genbank/DDBI databases.  
**RT**    ;  
**BT**    Botany, University of Texas at Austin, Austin, TX 78713, USA

**FH**    Key                  Location/Qualifiers  
**FT**    source             1..540  
         /db\_xref="taxon:3635"  
         /organism="Gossypium hirsutum"  
         /strain="Texas Marker-1"  
         /tissue\_type="24 days postanthesis secondary-stage fibers  
         /clone="CfI6S"  
         /cell\_type="seed epidermal cells"

**SQ**    Sequence 540 BP; 157 A; 105 C; 113 G; 165 T; 0 other;

Query Match                  100.0%; Score 10; DB 5; Length 540;  
Best Local Similarity      100.0%; Pred.No. 2.7e+04;  
Matches    10; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

Ov      1 TTTTGGAG 10  
         ||||| |  
         525 TTTTTGAAG 516

**RESULT**    2  
DZ0170/c                  DZ0170                  314 bp                  mRNA                  EST                  30-JUN-1996  
**LOCUS**    HUNGSO1144 Human promyelocyte Homo sapiens CDNA clone pml519 3'  
**DEFINITION**                  mRNA sequence.  
**DZ0170**                  DZ0170                  9301267  
                                 GI:501267  
**VERSION**                  DZ0170.1                  GI:501267  
**KEYWORDS**                  EST.  
**SOURCE**                  human.  
**ORGANISM**                  homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Euheria; Primates; Catarrhini; Hominiade; Homo.

**AUTHORS**                  1 (bases 1 to 314)  
Muraoka,K., Matsubara,K., Fukushima,A., Yoshii,T. and Okubo,K.  
Chromosomal assignments of 3'-directed partial cDNA sequences  
representing novel genes expressed in granulocytoid cells  
Genomics 23, 379-389 (1994)  
95137584  
2 (bases 1 to 314)  
Okubo,K., Fukushima,A., Yoshii,T., Miyama,T., Kojima,Y.,  
Yoshihari,H., Arimoto,J. and Matsubara,K.  
Gene expression of human promyelocytic cell line HL60 before and

after induction of differentiation. A new application of 3'-directed  
cDNA sequencing  
Unpublished (1993)

JOURNAL COMMENT

CONTACT: Okubo, K., Fukushima, A., Yoshii, J., Niijima, T., Kojima, Y.,  
Yoshinari, H., Arimoto, J. and Matsubara, K.  
Institute for Molecular and Cellular Biology  
Osaka University  
3-1 Yamada-oka, Suita, Osaka 565, Japan.  
Location/Qualifiers  
1. .314  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="6"  
/clone="pm1519"  
/clone\_lib="Human promyelocyte"  
/note="Female, adult, cell\_line = HL60, cell\_type =  
promyelocyte."

BASE COUNT 91 a 59 c 66 g 83 t 15 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 314;  
Best Local Similarity 100.0%; Pred. No. 3.3e+04;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTTTGGAG 10  
|||||||

DB 140 TTTTGGAG 131

RESULT 3  
D22028/ c 288 bp mRNA EST 20-JUL-1998  
LOCUS  
DEFINITION RICCI0206a Rice callus Oryza sativa cDNA clone R, mRNA sequence.  
ACCESSION D22028  
NID 9426169  
VERSION D22028.1 GI:426169  
KEYWORDS EST.  
SOURCE  
ORGANISM  
Oryza sativa.  
Oryza sativa.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
Poaceae; Oryza.  
1 (bases 1 to 288)  
Sasaki, T. and Minobe, Y.  
Rice cDNA from callus  
Unpublished (1994)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Contact: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program  
2-1-2 Kannondai, Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441  
Fax: 0298-38-7468  
Email: tsasaki@agr.affrc.go.jp  
PROJECT = 'RGP'

FEATURES  
SOURCE  
Location/Qualifiers  
1. .288  
/organism="Oryza sativa"  
/strain="Cultivar Nipponbare, sub\_species Japonica"  
/db\_xref="taxon:4530"  
/clone="R"  
/clone\_lib="Rice callus"  
/note="Vector: pBluescript II SK+; Site\_1: SalI; Site\_2:  
NotI; cDNA prepared from rice callus mRNAs by using  
oligo(dT) as a primer and ligating to the SalI-NotI site  
of pBluescript II SK+ phagemid.

BASE COUNT 78 a 57 c 59 g 92 t 2 others



Query Match 100.0%; Score 10; DB 20; Length 288;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTTTGGAG 10  
 |||||  
 Db 201 TTTTGGAG 192

RESULT 4  
 D22533 328 bp mRNA EST 20-JUL-1998  
 LOCUS D22533/c  
 DEFINITION RICC0190B Rice callus Oryza sativa cDNA clone R, mRNA sequence.  
 ACCESSION D22533  
 NID 9425855  
 VERSION D22533.1 GI:425855  
 KEYWORDS EST.  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
 Poaceae; Oryza.  
 1 (bases 1 to 328)  
 Sasaki, T. and Minobe, Y.  
 Rice cDNA from callus  
 Unpublished (1994)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

COMMENT  
 Contact: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program  
 2-1-2 Kannondai, Tsukuba  
 Ibaraki,  
 Japan 305  
 Tel: 0298-38-7441  
 Fax: 0298-38-7468  
 Email: tsasaki@abr.affrc.go.jp  
 PROJECT = "RGP",  
 Location/Qualifiers  
 1. .328  
 /organism="Oryza sativa"  
 /strain="cultivar Nipponbare, sub-species Japonica"  
 /db\_xref="taxon:4530"  
 /clone="R"  
 /clone\_1lb="Rice callus"  
 /note="Vector: pBluescript II SK+; Site\_1: SalI; Site\_2:  
 NotI; cDNA prepared from rice callus mRNAs by using  
 oligo(dT) as a primer and ligating to the SalI-NotI site  
 of pBluescript II SK+ phagemid.  
 96 a 69 c 81 g 78 t 4 others

3E COUNT  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTTTGGAG 10  
 |||||  
 Db 319 TTTTGGAG 310

RESULT 5  
 D23618 261 bp mRNA EST 20-JUL-1998  
 LOCUS D23618/c  
 DEFINITION RICC3103A Rice callus Oryza sativa cDNA clone R, mRNA sequence.  
 ACCESSION D23618  
 NID 9427550  
 VERSION D23618.1 GI:427550  
 KEYWORDS EST.  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
 Poaceae; Oryza.  
 1 (bases 1 to 261)  
 Sasaki, T. and Minobe, Y.  
 Rice cDNA from callus  
 Unpublished (1994)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

COMMENT  
 Contact: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program  
 2-1-2 Kannondai, Tsukuba  
 Ibaraki,  
 Japan 305  
 Tel: 0298-38-7441  
 Fax: 0298-38-7468  
 Email: tsasaki@abr.affrc.go.jp  
 PROJECT = "RGP",  
 Location/Qualifiers  
 1. .261  
 /organism="Oryza sativa"  
 /strain="cultivar Nipponbare, sub-species Japonica"  
 /db\_xref="taxon:4530"  
 /clone="R"  
 /clone\_1lb="Rice callus"  
 /note="Vector: pBluescript II SK+; Site\_1: SalI; Site\_2:  
 NotI; cDNA prepared from rice callus mRNAs by using  
 oligo(dT) as a primer and ligating to the SalI-NotI site  
 of pBluescript II SK+ phagemid.  
 79 a 49 c 52 g 79 t 2 others

BASE COUNT  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTTTGGAG 10  
 |||||  
 Db 251 TTTTGGAG 242

RESULT 6  
 D24868 331 bp mRNA EST 20-JUL-1998  
 LOCUS D24868/c  
 DEFINITION RICC2677A Rice root Oryza sativa cDNA clone R, mRNA sequence.  
 ACCESSION D24868  
 NID 9428776  
 VERSION D24868.1 GI:428716  
 KEYWORDS EST.  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
 Poaceae; Oryza.  
 1 (bases 1 to 331)  
 Minobe, Y. and Sasaki, T.  
 Rice cDNA from root  
 Unpublished (1995)

JOURNAL

COMMENT  
 Contact: Yuzo Minobe  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program  
 2-1-2 Kannondai, Tsukuba  
 Ibaraki,  
 Japan 305  
 Tel: 0298-38-7441  
 Fax: 0298-38-7468  
 Email: minobe@rcso.riken.go.jp  
 PROJECT = "RGP",  
 Location/Qualifiers  
 1. .331  
 /organism="Oryza sativa"  
 /strain="Nipponbare, sub-species Japonica"

/db\_xref="taxon:4530"  
 /clone="R"  
 /clone\_lib="Rice root"  
 /note="Prepared from seedling root."  
 BASE COUNT 89 a 63 c 68 g 109 t 2 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 331;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TTTTTCAG 10  
 Db 166 TTTTTCAG 157

FEATURES  
 922/c

DEFINITION D26922 423 bp mRNA EST 20-NOV-1995  
 CELK014D2R Yui Kohara unpublished cDNA Caenorhabditis elegans CDNA  
 clone YK14d2 3', mRNA sequence.

ACCESSION D26922  
 NID 9522008  
 VERSION D26922.1 GI:522008  
 KEYWORDS EST.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 423)  
 Kohara,Y., Mitsuki,H., Nishigaki,A., Motobashi,T., Sugimoto,A. and  
 Tabara,H.  
 TITLE Toward an expression map of the C.elegans genome  
 JOURNAL Unpublished (1994)  
 COMMENT

CONTACT: Yui Kohara  
 Gene Library Lab  
 National Institute of Genetics  
 Yata 111, Mishima, Shizuoka 411, Japan  
 Tel: 0559-75-0771  
 Fax: 0559-75-6240  
 Email: ykohara@dbj.nig.ac.jp.  
 Location/Qualifiers  
 1..423  
 /organism="Caenorhabditis elegans"  
 /strain="CB1489 him-8(e1489)"  
 /db\_xref="taxon:6239"  
 /clone="YK14d2"  
 /clone\_lib="Yui Kohara unpublished cDNA"  
 /note="dev-stage-varied, sex=Hermaphrodite male,  
 tissue=type-whole animal."  
 BASE COUNT 160 a 52 c 104 g 107 t  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 423;  
 Best Local Similarity 100.0%; Pred. No. 2.9e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TTTTTCAG 10  
 Db 404 TTTTTCAG 395

RESULT 8  
 D27177 398 bp mRNA EST 20-NOV-1995  
 LOCUS CELK004D5R Yui Kohara unpublished cDNA Caenorhabditis elegans CDNA  
 DEFINITION clone YK4d5 3', mRNA sequence.  
 D27177  
 ACCESSION 9521441  
 NID D27177.1 GI:521441  
 VERSION

KEYWORDS EST.  
 Caenorhabditis elegans.  
 SOURCE  
 ORGANISM Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 398)  
 Kohara,Y., Mitsuki,H., Nishigaki,A., Motobashi,T., Sugimoto,A. and  
 Tabara,H.  
 TITLE Toward an expression map of the C.elegans genome  
 JOURNAL Unpublished (1994)  
 COMMENT

CONTACT: Yui Kohara  
 Gene Library Lab  
 National Institute of Genetics  
 Yata 111, Mishima, Shizuoka 411, Japan  
 Tel: 0559-75-0771  
 Fax: 0559-75-6240  
 Email: ykohara@dbj.nig.ac.jp.  
 Location/Qualifiers  
 1..398  
 /organism="Caenorhabditis elegans"  
 /strain="CB1489 him-8(e1489)"  
 /db\_xref="taxon:6239"  
 /clone="YK4d5"  
 /clone\_lib="Yui Kohara unpublished cDNA"  
 /note="dev-stage-varied, sex=Hermaphrodite male,  
 tissue=type-whole animal."  
 BASE COUNT 162 a 42 c 88 g 105 t 1 others  
 ORIGIN

FEATURES  
 source

DEFINITION D27179 471 bp mRNA EST 20-NOV-1995  
 CELK002GXR Yui Kohara unpublished cDNA Caenorhabditis elegans CDNA  
 clone YK9g10 3', mRNA sequence.  
 D27179  
 ACCESSION D27179.1 GI:521341  
 NID 9521341  
 VERSION D27179.1 GI:521341  
 KEYWORDS EST.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 471)  
 Kohara,Y., Mitsuki,H., Nishigaki,A., Motobashi,T., Sugimoto,A. and  
 Tabara,H.  
 TITLE Toward an expression map of the C.elegans genome  
 JOURNAL Unpublished (1994)  
 COMMENT

Query Match 100.0%; Score 10; DB 20; Length 398;  
 Best Local Similarity 100.0%; Pred. No. 3e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TTTTTCAG 10  
 Db 230 TTTTTCAG 239

CONTACT: Yui Kohara  
 Gene Library Lab  
 National Institute of Genetics  
 Yata 111, Mishima, Shizuoka 411, Japan  
 Tel: 0559-75-0771  
 Fax: 0559-75-6240  
 Email: ykohara@dbj.nig.ac.jp.  
 Location/Qualifiers  
 1..471  
 /organism="Caenorhabditis elegans"  
 /strain="CB1489 him-8(e1489)"  
 /db\_xref="taxon:6239"  
 /clone="YK2g10"  
 /clone\_lib="Yui Kohara unpublished cDNA"  
 BASE COUNT 160 a 52 c 104 g 107 t  
 ORIGIN

DEFINITION D27179 471 bp mRNA EST 20-NOV-1995  
 CELK002GXR Yui Kohara unpublished cDNA Caenorhabditis elegans CDNA  
 clone YK9g10 3', mRNA sequence.  
 D27179  
 ACCESSION D27179.1 GI:521341  
 NID 9521341  
 VERSION D27179.1 GI:521341  
 KEYWORDS EST.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 471)  
 Kohara,Y., Mitsuki,H., Nishigaki,A., Motobashi,T., Sugimoto,A. and  
 Tabara,H.  
 TITLE Toward an expression map of the C.elegans genome  
 JOURNAL Unpublished (1994)  
 COMMENT

CONTACT: Yui Kohara  
 Gene Library Lab  
 National Institute of Genetics  
 Yata 111, Mishima, Shizuoka 411, Japan  
 Tel: 0559-75-0771  
 Fax: 0559-75-6240  
 Email: ykohara@dbj.nig.ac.jp.  
 Location/Qualifiers  
 1..471  
 /organism="Caenorhabditis elegans"  
 /strain="CB1489 him-8(e1489)"  
 /db\_xref="taxon:6239"  
 /clone="YK2g10"  
 /clone\_lib="Yui Kohara unpublished cDNA"  
 BASE COUNT 160 a 52 c 104 g 107 t  
 ORIGIN

FEATURES  
 source

DEFINITION D27179 471 bp mRNA EST 20-NOV-1995  
 CELK002GXR Yui Kohara unpublished cDNA Caenorhabditis elegans CDNA  
 clone YK9g10 3', mRNA sequence.  
 D27179  
 ACCESSION D27179.1 GI:521341  
 NID 9521341  
 VERSION D27179.1 GI:521341  
 KEYWORDS EST.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 471)  
 Kohara,Y., Mitsuki,H., Nishigaki,A., Motobashi,T., Sugimoto,A. and  
 Tabara,H.  
 TITLE Toward an expression map of the C.elegans genome  
 JOURNAL Unpublished (1994)  
 COMMENT

BASE COUNT 184 a 70 c 99 g 118 t  
 ORIGIN /note="dev-stage=varied, sex=Hermaphrodite male, tissue-type=whole animal"

Query Match 100.0%; Score 10; DB 20; Length 471;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTTGAAG 10  
 |||||  
 Db 98 TTTTGAAG 107

RESULT 10

D27520 461 bp mRNA EST 20-NOV-1995  
 LOCUS CELK012B4R Yuj1 Kohara unpublished CDNA Caenorhabditis elegans CDNA  
 DEFINITION clone yk12b4 3', mRNA sequence.  
 D27520

ISSION 9523006  
 VERSION D27520.1 GI:523006  
 KEYWORDS EST.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.

REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 461)  
 AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and  
 Tabara,H.  
 TITLE Toward an expression map of the C.elegans genome  
 JOURNAL Unpublished (1994)  
 COMMENT

CONTACT: Yuj1 Kohara  
 Gene Library Lab  
 National Institute of Genetics  
 Yata 1111, Mishima, Shizuoka 411, Japan  
 Tel: 0559-75-0771  
 Fax: 0559-75-6240  
 Email: ykohara@dbj.nig.ac.jp.  
 Location/Qualifiers

FEATURES

source  
 1..461  
 /organism="Caenorhabditis elegans"  
 /strain="CB1489 him-8(e1489)"  
 /db\_xref="taxon:6239"  
 /clone="yk12b4"  
 /clone\_id="Yuj1 Kohara unpublished CDNA"  
 /note="dev-stage=varied, sex=Hermaphrodite male, tissue-type=whole animal"  
 BASE COUNT 144 a 73 c 89 g 155 t  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 461;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTTGAAG 10  
 |||||  
 Db 226 TTTTGAAG 235

RESULT 11

D27544 417 bp mRNA EST 20-NOV-1995  
 LOCUS CELK013A7R Yuj1 Kohara unpublished CDNA Caenorhabditis elegans CDNA  
 DEFINITION clone yk13a7 3', mRNA sequence.  
 D27544

ACCESSION 9523039  
 NID D27544.1 GI:523039  
 VERSION EST.  
 KEYWORDS Caenorhabditis elegans.

ORGANISM

Caenorhabditis elegans

REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 417)  
 AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and  
 Tabara,H.  
 TITLE Toward an expression map of the C.elegans genome  
 JOURNAL Unpublished (1994)  
 COMMENT

CONTACT: Yuj1 Kohara  
 Gene Library Lab  
 National Institute of Genetics  
 Yata 1111, Mishima, Shizuoka 411, Japan  
 Tel: 0559-75-0771  
 Fax: 0559-75-6240  
 Email: ykohara@dbj.nig.ac.jp.  
 Location/Qualifiers

FEATURES

source  
 1..417  
 /organism="Caenorhabditis elegans"  
 /strain="CB1489 him-8(e1489)"  
 /db\_xref="taxon:6239"  
 /clone="yk13a7"  
 /clone\_id="Yuj1 Kohara unpublished CDNA"  
 /note="dev-stage=varied, sex=Hermaphrodite male, tissue-type=whole animal"  
 BASE COUNT 112 a 85 c 75 g 144 t 1 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 417;  
 Best Local Similarity 100.0%; Pred. No. 2.9e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTTGAAG 10  
 |||||  
 Db 296 TTTTGAAG 287

RESULT 12

D27856 403 bp mRNA EST 20-NOV-1995  
 LOCUS CELK005B2R Yuj1 Kohara unpublished CDNA Caenorhabditis elegans CDNA  
 DEFINITION clone yk5b2 3', mRNA sequence.  
 D27856

ACCESSION 9522587  
 NID D27856.1 GI:522587  
 VERSION EST.  
 KEYWORDS Caenorhabditis elegans.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.

REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 403)  
 AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and  
 Tabara,H.  
 TITLE Toward an expression map of the C.elegans genome  
 JOURNAL Unpublished (1994)  
 COMMENT

CONTACT: Yuj1 Kohara  
 Gene Library Lab  
 National Institute of Genetics  
 Yata 1111, Mishima, Shizuoka 411, Japan  
 Tel: 0559-75-0771  
 Fax: 0559-75-6240  
 Email: ykohara@dbj.nig.ac.jp.  
 Location/Qualifiers

FEATURES

source  
 1..403  
 /organism="Caenorhabditis elegans"  
 /strain="CB1489 him-8(e1489)"  
 /db\_xref="taxon:6239"  
 /clone="yk5b2"  
 /clone\_id="Yuj1 Kohara unpublished CDNA"  
 /note="dev-stage=varied, sex=Hermaphrodite male, tissue-type=whole animal"



Query Match 100.0%; Score 10; DB 20; Length 403;  
Best Local Similarity 100.0%; Pred. No. 3e+04;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 TTTTGAAG 10  
|||||  
Db 69 TTTTGAAG 60

Search completed: September 17, 1999, 21:27:56  
Job time: 14260 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:33:26; Search time 425.19 Seconds

(without alignments)  
5.884 Million cell updates/sec

Title: US-08-956-518a-85

Perfect score: 10

Sequence: 1 TCTCCTTAG 10

Scoring table: IDENTITY\_NUC

Archived: 311585 seqs, 125096042 residues

--cbase: N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	100.0	1476	1 N90613	CD20 CDNA. Rapid 1
2	10	100.0	1473	1 N90610	CD20.4 CDNA. Rapid
3	10	100.0	1982	1 N90225	Malaria-specific P
4	10	100.0	5376	1 N90732	Sequence of plasmid
5	10	100.0	1589	1 N90733	Genomic DNA from P
6	10	100.0	1426	1 N91650	DNA encoding Sn-gi
7	10	100.0	1254	1 N92837	CDNA from Plasmid
8	10	100.0	1717	1 N04266	Probe containing c
9	10	100.0	1764	1 N81290	New recombinant CD
10	10	100.0	1791	1 N05796	Encodes B. t. t inse
11	10	100.0	102	1 N71032	Sequence of synthe
12	10	100.0	3744	1 Q10157	Sequence contg. ab
13	10	100.0	1470	1 Q10869	Plasmidum vivax 6
14	10	100.0	1121	1 Q10245	Clone 9C encoding
15	10	100.0	906	1 Q10246	Encodes urate oxid
16	10	100.0	906	1 Q10247	Urate oxidase-enco
17	10	100.0	920	1 Q10248	Urate oxidase-enco
18	10	100.0	980	1 Q10325	Promoter of nodAB
19	10	100.0	1014	1 Q10834	Encodes kappa 11gn
20	10	100.0	1121	1 Q12430	Urate oxidase codl
21	10	100.0	1013	1 Q12431	Fragment D of urat
22	10	100.0	2781	1 Q13339	HSF CDNA sequence.
23	10	100.0	102	1 N50151	Sequence of thymos
24	10	100.0	98	1 N50083	Thymosin alpha 1 s
25	10	100.0	102	1 N50386	Sequence of synthe
26	10	100.0	1260	1 Q20265	Protein D - human
27	10	100.0	1764	1 Q20794	Newcastle Disease
28	10	100.0	882	1 Q21316	DNA encoding envel
29	10	100.0	1474	1 Q21170	Clone CD20.4 encod
30	10	100.0	1476	1 Q21173	Human CD20 antigen
31	10	100.0	906	1 Q22321	Sequence of plasmid
32	10	100.0	56	1 Q22759	Sequence of oligos
33	10	100.0	102	1 Q27178	Sequence of thymos
34	10	100.0	102	1 N30067	Adrenodoxin reduct
35	10	100.0	5089	1 Q27764	Sequence of synthe
36	10	100.0	102	1 N10021	Sequence of synthe
37	10	100.0	2781	1 Q25712	Partial exotoxin A
38	10	100.0	201	1 Q28261	Odorant receptor C
39	10	100.0	643	1 Q29865	Korean hepatitis C
40	10	100.0	9472	1 Q33882	KHCv 937 probe P3
41	10	100.0	30	1 Q33118	Expressed Sequence
42	10	100.0	440	1 Q39849	Expressed Sequence
43	10	100.0	376	1 Q39878	Expressed Sequence

## ALIGNMENTS

C 44 10 100.0 163 1 Q39914 Expressed Sequence  
C 45 10 100.0 260 1 X27003 Differentially exp

RESULT 1  
ID N90613 standard; CDNA; 1476 BP.  
AC N90613;  
DT 20-DEC-1989 (first entry).  
DE CD20 CDNA.  
KW Tyrosine suppressor gene; ASV LTR; human cytomegalovirus AD169 enhancer;  
KW HIV box; immunoselection; immune deficiency diseases; vasculitis;  
KW systemic lupus erythematosus; rheumatoid arthritis; neoplasms.  
PN EP-330191-A.  
PD 30-AUG-1989.  
PF 23-FEB-1989; 103127.  
PR 25-FEB-1989; US-160416.  
PA (GEO) Gen. Hospital Corp.  
PI Seed B, Allen J, Aruffo A, Camerini D, Lauffer L, Oquendo CP,  
PI Simmons D, Stamenkovic I, Stengelin S;  
DR WPI; 89-250302/35.  
PT Rapid immuno-selection cloning - used to clone genes encoding  
PT cell surface antigens associated with mammalian T lymphocytes.  
PS Disclosure; fig. 13; 69pp; English.  
CC CD20 encodes a cell surface antigen involved in cell mediated immunity.  
CC This DNA can be expressed in a vector which transforms COS cells. The  
CC vector can isolate any protein, and clones are easy to manipulate.  
SQ Sequence 1476 BP; 435 A; 324 C; 285 G; 432 T;

Query Match 100.0%; Score 10; DB 1; Length 1476;  
Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAG 10  
DB 981 TCTCCTTAG 990

RESULT 2  
ID N90610 standard; CDNA; 1473 BP.  
AC N90610;  
DT 20-DEC-1989 (first entry).  
DE CD20.4 CDNA.  
KW Tyrosine suppressor gene; ASV LTR; human cytomegalovirus AD169 enhancer;  
KW HIV box; immunoselection; immune deficiency diseases; vasculitis;  
KW systemic lupus erythematosus; rheumatoid arthritis; neoplasms.  
FH Key Location/Qualifiers  
FT cds 94..987  
FT /\*tag= a  
FT /\*tag= b  
FT /\*tag= b  
FT /\*note= site of polyA+ tail in CD20.6  
PN EP-330191-A.  
PD 30-AUG-1989.  
PF 23-FEB-1989; 103127.  
PR 25-FEB-1989; US-160416.  
PA (GEO) Gen. Hospital Corp.  
PI Seed B, Allen J, Aruffo A, Camerini D, Lauffer L, Oquendo CP,  
PI Simmons D, Stamenkovic I, Stengelin S;  
DR WPI; 89-250302/35.  
DR P-PSDB; P91356.  
PT Rapid immuno-selection cloning - used to clone genes encoding  
PT cell surface antigens associated with mammalian T lymphocytes.  
PS Disclosure; fig. 10; 69pp; English.  
CC CD20.4 encodes a cell surface antigen involved in cell mediated immunity.  
CC This DNA can be expressed in a vector which transforms COS cells. The  
CC vector can isolate any protein, and clones are easy to manipulate.  
SQ Sequence 1473 BP; 434 A; 321 C; 286 G; 432 T;

Query Match 100.0%; Score 10; DB 1; Length 1473;  
 Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCTCCTTAG 10  
 Db 979 TCTCCTTAG 988

RESULT 3  
 N90225/C  
 ID N90225 standard; DNA; 1982 BP.  
 AC N90225;  
 DT 1-NOV-1989 (first entry)  
 DE Malaria-specific Plasmodium falciparum protein.  
 KM DNA: malaria; 41kd protein; Plasmodium falciparum; vaccine.  
 OS Plasmodium falciparum  
 Fu Key Location/Qualifiers  
 cds 526..1611  
 /\*tag- a

PN EP-322712-A.  
 PD 05-JUL-1989.  
 PF 20-DEC-1988; 121299.  
 PR 30-DEC-1987; DE-831351.  
 PA (BEHW) Behringwerke.  
 PI Knapp B, Hundt E, Enders B, Kupper H;  
 DR WPI: 89-194071/27.  
 P-PSDB: P90419.  
 PT New antigenic proteins from Plasmodium falciparum - new encoding nucleic acid sequences and derived antibodies, useful in vaccines, diagnosis etc.  
 PS Claim 1; page 23 and Table 18; 25pp; German.  
 CC The DNA encodes a malaria-specific P. falciparum 41kd protein (see P90419) is useful in a protective vaccine, esp. against malaria. Produced antibodies are useful for passive immunisation, and antibodies, the protein and the DNA sequence are useful as diagnostic reagents. See also N90211-25 and P90403-19.  
 CC Sequence 1982 BP; 764 A; 230 C; 287 G; 701 T;  
 SQ

Query Match 100.0%; Score 10; DB 1; Length 1982;  
 Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCTCCTTAG 10  
 Db 1302 TCTCCTTAG 1293

RESULT 4  
 N90732/C  
 ID N90732 standard; DNA; 5376 BP.  
 AC N90732;  
 DT 18-JUN-1990 (first entry)  
 DE Sequence of plasmid p8/3  
 KM Plasmid p8/3; vector; Plasmodium falciparum merozoite antigen.  
 FH Key Location/Qualifiers  
 FT cds 115..1257  
 /\*tag- a  
 /note-"It codes for a 41kd polypeptide"  
 FT signal\_peptide 115..168  
 /\*tag- b  
 FT mat\_peptide 169..1257  
 /\*tag- c  
 EP-309746-A.  
 PD 05-APR-1989.  
 PF 27-AUG-1988; 114016.  
 PR 08-SEP-1987; CH-003486.  
 PA (HOFF) Hoffmann-La Roche AG.  
 PI Cert a U;  
 DR WPI: 89-101095/14.  
 PT New peptide(s) -

PT contg. epitope(s) of Plasmodium falciparum merozoite antigen,  
 for use in malaria vaccines  
 PS Figure 11a-11e; 6/PP; German.  
 CC Plasmid p8/3 is used as an expression vector.  
 SQ Sequence 5376 BP; 1592 A; 1152 C; 1213 G; 1419 T;

Query Match 100.0%; Score 10; DB 1; Length 5376;  
 Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCTCCTTAG 10  
 Db 945 TCTCCTTAG 936

RESULT 5  
 N90733/C  
 ID N90733 standard; DNA; 1599 BP.  
 AC N90733;  
 DT 18-JUN-1990 (first entry)  
 DE Genomic DNA from Plasmodium falciparum KI isolates encoding the 41 kd merozoite antigen  
 KM Plasmodium falciparum KI isolate; merozoite antigen; epitope; immunogen;  
 KM anti-malaria vaccine; passive immunisation; malaria diagnosis.  
 OS Plasmodium falciparum  
 Fu Key Location/Qualifiers  
 FT cds 346..1434  
 /\*tag- a  
 FT conflict 619..622  
 /\*tag- b  
 /\*note-"Differs from cDNA from M25 isolate of P. falciparum."  
 FT conflict 718..721  
 /\*tag- c  
 /\*note-"Differs from cDNA from M25 isolate of P. falciparum"  
 FT conflict 949..952  
 /\*tag- d  
 /\*note-"Differs from cDNA from M25 isolate of P. falciparum"

PN EP-309746-A.  
 PD 05-APR-1989.  
 PF 27-AUG-1988; 114016.  
 PR 08-SEP-1987; CH-003486.  
 PA (HOFF) Hoffmann-La Roche AG.  
 PI Cert a U;  
 DR WPI: 89-101095/14.  
 P-PSDB: P93566.  
 PT New peptide(s) -  
 PT contg. epitope(s) of Plasmodium falciparum merozoite antigen,  
 for use in malaria vaccines  
 PS Figure 12a-12d; 67pp; German.  
 CC The genomic DNA of isolate RO-33 Ghana was largely identical to this in 3 codons (see n92937). Plasmodium falciparum merozoite antigen can be coupled to an affinity peptide, or adsorbed or covalently bonded on a carrier. It, and its derivatives, are useful as immunogens in anti-malaria vaccines. Antibodies directed against it are useful for passive immunisation and diagnosis.  
 CC Sequence 1599 BP; 619 A; 207 C; 259 G; 514 T;  
 SQ

Query Match 100.0%; Score 10; DB 1; Length 1599;  
 Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCTCCTTAG 10  
 Db 1122 TCTCCTTAG 1113

RESULT 6

```

CC (see n90733). Plasmodium falciparum merozoite antigen can be
CC coupled to an affinity peptide, or adsorbed or covalently
CC bonded on a carrier. It, and its derivatives, are useful as immunogens in
CC anti-malaria vaccines. Antibodies directed against it are useful for
CC passive immunisation and diagnosis.
SQ Sequence 1254 BP; 471 A; 186 C; 229 G; 368 T;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 1554;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY 1 TCTCCTTAAG 10
|||||
DB 813 TCTCCTTAAG 804

RESULT 8
Q04266/c
ID Q04266 standard; cDNA; 1717 BP.
AC Q04266;
DT 13-SEP-1990 (first entry)
DE Probe containing coding exons for murine Ig light chain J regions
KW chimeric antibodies; TAG72; variable region of immunoglobulin light chain;
KW variable region of immunoglobulin heavy chain; light chain J region.
OS synthetic.
PN EP-365997-A.
PD 02-MAY-1990.
PF 18-OCT-1989; 119361.
PR 19-OCT-1988; US-259943.
PA (DOWC) Dow Chemical Co.
PI Mezes P, Gourlie B, Rixon M;
DR WPI: 90-133521/18.
PT Chimeric antibodies against TAG72 and conjugate to provide imaging
PT markers and therapeutic tools
PS Disclosure; pp; English.
CC The sequence is the 1.71kb HindIII-PstI fragment isolated from plasmid
CC pcg1. It contains the coding exons for murine light chain J regions i.e.
CC J1-J5. The probe was used to isolate the CC49 light chain for inclusion
CC in chimeric antibodies. A typical chimera is composed of
CC variable regions of heavy and light chains with high affinity for TAG72,
CC (a tumour-associated glycoprotein) and constant regions from human
CC antibodies. It is thought that these chimeric antibodies have fewer side-
CC effects when administered to human patients due to the constant regions
CC being of human origin.
CC See also Q04259-Q04265 and Q04267-9.
SQ Sequence 1717 BP; 492 A; 297 C; 425 G; 503 T;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 1717;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY 1 TCTCCTTAAG 10
|||||
DB 554 TCTCCTTAAG 545

RESULT 9
N81290/c
ID N81290 standard; cDNA; 1764 BP.
AC N81290;
DT 22-OCT-1990 (first entry)
DE New recombinant cDNA coding for the fusion protein F0 of Newcastle
DE disease virus (NDV)
KW vaccines; poultry; clone; II.14(F); Newcastle disease virus; protein F0;
KW fowl vaccine; ss.
OS Newcastle disease virus.
FH Key location/Qualifiers
FT signal_peptide 68..121
FT /*tag- a
FT /*tag- b
FT /*tag- b
FT /*tag- b
FT /*tag- b

```





PR 02-JUN-1989; US-360294.  
 PA (UYGE-) UNIV GEORGIA RES FO.  
 PI Timberlake W, Adams TE, Miralito PM;  
 DR WPI; 91-007214/01.  
 DE R-PSDB; R10047.  
 CC DNA sequence of abaa gene of *Aspergillus nidulans* and gene fusion  
 PT - with inducible promoter for control of growth and  
 PS differentiation of filamentous fungi  
 PS Disclosure; page 25; 36pp; English.  
 CC This DNA fragment encodes the *A.nidulans* abacus (abaa) gene follow-  
 CC ing processing to remove the 2 introns. The protein encoded is  
 CC capable of regulating growth and development in filamentous fungi.  
 CC It is pref. fused to an inducible promoter, e.g. nad or prnA,  
 CC enabling the mis-scheduled expression of the abaa gene, inducing  
 CC abortive development in the fungus. This enables increased prodn.  
 CC of medically or commercially important substances which are normally  
 CC only produced by the fungi during conidiation.  
 CC Sequence 3744 BP; 933 A; 1100 C; 793 G; 918 T;

Query Match 100.0%; Score 10; DB 1; Length 3744;  
 Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCTCCTTAG 10  
 |||||  
 DB 176 TCTCCTTAG 185

RESULT 13  
 Q10869/c  
 ID Q10869 standard; DNA; 1470 BP.  
 AC Q10869;  
 DT 15-APR-1991 (first entry)  
 DE Plasmodium vivax 66kD merozoite antigen gene partial sequence.  
 KW Plasmodium merozoite antigens; malaria vaccine; ss.  
 OS Plasmodium vivax.  
 FH Key  
 FT cds  
 FT 1. 1470  
 FT /tag= a  
 FT /product= part of P.vivax merozoite antigen protein  
 PN US7483516-A.  
 PD 13-JAN-1991;  
 PF 22-FEB-1990; 483516.  
 PR (USSH ) NAT INST OF HEALTH.  
 PR Waters AP, McCutchan TF;  
 WPI: 91-044381/06.  
 DR P-PSDB; R10935.  
 CC Plasmodium merozoite antigen proteins and DNA sequences - useful  
 CC in vaccine prodn., anti-malarial drug design, and in diagnostics  
 CC Disclosure; fig 1; 31pp; English.  
 CC This sequence is contained in a recombinant DNA molecule com-  
 CC prising e.g. a pUC19 or vaccinia virus vector sequence and  
 CC regulatory elements. This allows efficient expression of the  
 CC antigen on transformation of host cells. The P.vivax antigen  
 CC prodn. is useful in an anti-malarial vaccine for conferring immu-  
 CC nity against the merozoite form of the parasite and preventing in-  
 CC fection of uninfected red blood cells. See also Q10418.  
 CC Sequence 1470 BP; 496 A; 279 C; 361 G; 334 T;

Query Match 100.0%; Score 10; DB 1; Length 1470;  
 Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCTCCTTAG 10  
 |||||  
 DB 214 TCTCCTTAG 205

RESULT 14  
 Q10245/c

ID Q10245 standard; cDNA; 1121 BP.  
 AC Q10245;  
 DT 27-MAR-1991 (first entry)  
 DE Clone 9C encoding a protein with urate oxidase activity.  
 KW urate oxidase; uricase; hyperuricaemia; kidney stones; chemotherapy;  
 KW ss.  
 OS *Aspergillus flavus*.  
 FH Key  
 FT cds  
 FT Location/Qualifiers  
 FT 109..1017  
 FT /tag= a  
 FT /product= urate oxidase-like protein

EP-408461-A.  
 PD 16-JAN-1991.  
 PF 13-JUL-1990; 402023.  
 PR 13-JUL-1989; FR-009550.  
 PR 29-DEC-1989; FR-017466.  
 PR 06-FEB-1990; FR-001368.  
 PA (SNFI ) SANOFI SA.  
 PI Caput D, Ferrara P, Guillemot JC, Kaghad M, Legoux R;  
 PI Loison G, Labree, Lupker J;  
 DR WPI: 91-016644/03.  
 DE P-PSDB; R10222.  
 CC New protein with high urate oxidase activity - and recombinant  
 CC DNA encoding it, vectors and transformed cells, used for treating  
 CC hyperuricaemia, etc.  
 PS Claim 8; Fig 3; 68pp; French.  
 CC mRNA was isolated from a urate oxidase-producing strain of *A.flavus*  
 CC and reverse transcribed into cDNA. A gene bank was constructed from  
 CC *E.coli* MC 1061 containing plasmid pTZ19R containing the cDNA. Probes  
 CC were synthesised based upon the amino acid sequence of A.flavus  
 CC urate oxidase and these used to screen the gene bank. Five positive  
 CC colonies were analysed; this is the sequence of the longest cDNA  
 CC insert ("clone 9C"). See also Q10246-Q10253, Q10255.  
 CC Sequence 1121 BP; 281 A; 333 C; 264 G; 243 T;

Query Match 100.0%; Score 10; DB 1; Length 1121;  
 Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCTCCTTAG 10  
 |||||  
 DB 628 TCTCCTTAG 619

RESULT 15  
 Q10246/c  
 ID Q10246 standard; DNA; 906 BP.  
 AC Q10246;  
 DT 27-MAR-1991 (first entry)  
 DE Encodes urate oxidase specifically for expression in prokaryotes.  
 KW urate oxidase; uricase; hyperuricaemia; kidney stones; chemotherapy;  
 KW ss.  
 OS Synthetic.  
 FH Key  
 FT cds  
 FT Location/Qualifiers  
 FT 1..906  
 FT /tag= a  
 FT /product= urate oxidase  
 PN EP-408461-A.  
 PD 16-JAN-1991.  
 PF 13-JUL-1990; 402023.  
 PR 13-JUL-1989; FR-009550.  
 PR 29-DEC-1989; FR-017466.  
 PR 06-FEB-1990; FR-001368.  
 PA (SNFI ) SANOFI SA.  
 PI Caput D, Ferrara P, Guillemot JC, Kaghad M, Legoux R;  
 PI Loison G, Labree, Lupker J;  
 DR WPI: 91-016644/03.  
 DE P-PSDB; R10222.  
 CC New protein with high urate oxidase activity - and recombinant  
 CC DNA encoding it, vectors and transformed cells, used for treating  
 CC hyperuricaemia, etc.  
 PS Claim 10; Page 50; 68pp; French.

CC mRNA was isolated from a urate oxidase-producing strain of *A. flavus*  
CC and reverse transcribed into cDNA. A gene bank was constructed from  
CC E.coli MC 1061 containing plasmid pTZ19R containing the cDNA. Probes  
CC were synthesised based upon the amino acid sequence of *A. flavus*  
CC urate oxidase and these used to screen the gene bank. Five positive  
CC clones were analysed, one of which contained inserts of 1.2kb (-  
CC "clone 9c"). An *AccI*-*XbaI* fragment was isolated from 9c and ligated  
CC to a synthetic fragment designed to replace codons immediately  
CC downstream of the start ATG with codons typical of prokaryotic  
CC genes (substituted bases are found at positions 6, 9, 21 and 42).  
CC The recombinant sequence was ligated to plasmid p462 to give p466.  
CC See also Q10245 and Q10247-Q10253 and Q10255.  
SQ Sequence 906 BP; 219 A; 267 C; 241 G; 179 T;

Query Match 100.0%; Score 10; DB 1; Length 906;  
Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0" 1 TCTCCTTAG 10  
|||||||  
520 TCTCCTTAG 511

Search completed: September 18, 1999, 00:33:28  
Job time: 18968 sec

Result No.	Score	Query Match	Length	DB	ID	Description
1	7	12.7	9360	1	002830	CDNA to HIV-2 RNA.
2	7	12.7	2809	1	002842	PKS11 Plasmid car
3	7	12.7	2026	1	002894	CDNA insert of vec
4	7	12.7	117	1	003306	Gene encoding prot
5	7	12.7	558	1	003312	RAP2 gene encoding
6	7	12.7	1766	1	003247	Aspergillopepsin A
7	7	12.7	2122	1	003216	SalI restriction f
8	7	12.7	3168	1	003345	pelc Pectin lyase
9	7	12.7	2774	1	003345	peib Pectin lyase
10	7	12.7	1763	1	003332	Gene encodes functi
11	7	12.7	354	1	003335	pBR32 secretion exp
12	7	12.7	2537	1	002814	Sequence of pre-Tg
13	7	12.7	2676	1	002819	CDNA sequence enco
14	7	12.7	1389	1	003218	DNA sequence enco
15	7	12.7	1383	1	001794	Nucleotide sequenc
16	7	12.7	1873	1	001648	Buf-3 gene for hum
17	7	12.7	1084	1	003671	Zucchini yellow mo
18	7	12.7	2967	1	003659	Maize C1 sequence
19	7	12.7	2546	1	003661	Maize C3 sequence
20	7	12.7	2038	1	003663	Fragment of tobacco
21	7	12.7	1392	1	003542	Pro-type human pla
22	7	12.7	1392	1	003542	Pro-type human pla
23	7	12.7	2245	1	003543	Chromosomal gene f
24	7	12.7	2245	1	003543	Chromosomal gene f
25	7	12.7	1661	1	003129	DNA homologous to
26	7	12.7	2260	1	003118	HindIII/HpaI frag
27	7	12.7	1368	1	001865	Nucleotide sequenc
28	7	12.7	1185	1	001862	Nucleotide sequenc
29	7	12.7	1567	1	001863	Sequence of the rI
30	7	12.7	774	1	003621	19Kd maize c219a2
31	7	12.7	553	1	003735	MPB-57 protein enc
32	7	12.7	3926	1	003736	Sequence of plasmid
33	7	12.7	7368	1	004013	Acetyl-CoA-carboxy
34	7	12.7	7368	1	004013	Acetyl-CoA-carboxy
35	7	12.7	337	1	003261	Heat-resistant RNA
36	7	12.7	1017	1	003261	Cephalosporium acr
37	7	12.7	4897	1	003259	Pseudotuberculosis
38	7	12.7	4897	1	003259	Pseudotuberculosis
39	7	12.7	488	1	003266	Phosphonothricin-r
40	7	12.7	287	1	003824	Translation activat
41	7	12.7	5400	1	003861	Thymidine kinase
42	7	12.7	5400	1	003861	Thymidine kinase
43	7	12.7	3545	1	003817	Marok's disease vi

```

c   44      7      12.7    2040    1    003810      GC gene of R81B st
c   45      7      12.7    2036    1    X26936      CDNA encoding a hu

                                ALIGNMENTS

RESULT      1
002830
ID          002830 standard; cDNA; 9360 BP.
AC          Q02830;
DT          29-MAY-1989 (first entry)
DE          CDNA to HIV-2 RNA.
KW          HIV, AIDS; Vaccine; pUC-HIV-2(GH-1).
OS          HIV-2.
FH          Key
EH          Location/Qualifiers
FI          cds
FT          /tag= a
              544..2109
              //label=Cag reading frame
              1830..4934
              /tag= b
              //label=Pol gene reading frame.
              4867..5511
              /tag= c
              //label=G gene reading frame
              5342..5677
              /tag= d
              //label=X gene reading frame
              5680..5994
              /tag= e
              //label=R gene reading frame
              6148..8703
              /tag= f
              //label=Env gene reading frame.
              8540..9304
              /tag= g
              //label=F gene reading frame.

FN          J01289486-A.
PN          21-NOV-1989.
PD          16-MAY-1988; 119024.
PE          16-MAY-1988; JP-119024.
PR          Toa Nentryo kogyo KK, Fuji Kbio KK.
PA          WPI; 90-005177/01.
DR          P-PSDB; R04024-30.
DT          DNA indicating complement to RNA gene -
PT          PT of Human Immunodeficiency Virus type 2 used for new vaccine or
PF          diagnostic for AIDS virus.
PP          Claim 2; Fig. 4; 12pp; Japanese.
PS          CDNA to novel HIV-2 (GH-1) has been integrated into plasmid
SC          pUC HIV-2(GH-1). Useful for diagnosis and vaccination against the virus.
CC          Carries 7 overlapping genes in various reading frames, including gag,
CG          Pol and Env.
CO          Sequence 9360 BP; 3211 A; 1944 C; 2317 G; 1888 T;

Query Match      12.7%; Score 7; DB 1; Length 9360;
Best Local Similarity 100.0%; Pred. No. 6.9e+03;
Matches       7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY          49 CTGCACG 55
            |||||
DB          1086 CTGCACG 1092

RESULT      2
002842/c
ID          002842 standard; DNA; 2809 BP.
AC          Q02842;
DT          01-JUN-1990 (first entry)
DE          PKS311 Plasmid carrying mature fibrolase.
KW          Crocalidus; fibrolase; thrombolytic; ds.
AG          Agriastrodon conforrix.
FH          Key
EH          Location/Qualifiers
```

FT cds 405.1838  
/\*tag- a  
PN ZA8809415-A.  
PD 27-SEP-1989.  
PF 15-DEC-1988; 889415.  
PR 18-DEC-1987; US-124981.  
PA (CHIR) Chiron Corp.  
PI Valenzuela P, Brake A, Randolph A;  
DR WPI: 90-007677/01.  
PT Crocalidus venom fibrolyase prodn. -  
PS also yeast expression vectors contg. DNA and recombinant DNA.  
CC Example 5; Page 29; 64pp; English.  
CC PK3311 plasmid as carried by the pAB24 yeast vector with GAP promoter.  
CC alpha factor leader (S.cerevisiae) encoding mature fibrolyase.  
CC Fibrolyases are thrombolytic agents, useful in treating thromboembolic  
CC conditions e.g. pulmonary embolism.  
CC See also Q02837 and Q02838.  
SQ Sequence 2809 BP; 887 A; 552 C; 538 G; 832 T;  
Query Match 12.7%; Score 7; DB 1; Length 2809;  
Best Local Similarity 100.0%; Pred. No. 7.6e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 49 CTGCACG 55  
DB 2286 CTGCACG 2280  
RESULT 3  
002894  
ID Q02894 standard; DNA: 2026 BP.  
AC Q02894;  
DT 11-JUL-1990 (first entry)  
DE cDNA insert of vector PSP 6K-13-7 encoding pulmonary surfactant protein.  
KW Respiratory distress syndrome; RDS; hyaline membrane  
KW disease; pulmonary surfactant protein.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT cds 14.1156  
/\*tag- a  
FT mat\_peptide 614.811  
/\*tag- b  
PN US482422-A.  
PD 21-NOV-1989.  
PF 24-SEP-1987; 100372.  
PR 26-SEP-1985; US-791120.  
PR 15-AUG-1986; US-897183.  
(GENE-) Genetics Institute.  
TA Taensh HW, Jacobs KA, Steinbrink DR, Floros J, Phelps DS, Fritsch EF;  
DR WPI: 90-036829/05.  
PT Purified human pulmonary surfactant protein -  
PT useful for treating respiratory distress syndrome.  
PS Disclosure: Table 6; 15pp; English.  
CC Protein is useful in treatment of respiratory distress syndrome (Hyaline  
CC membrane disease) enhancing pulmonary surfactant activity.  
CC Protein is encoded by the cDNA insert in vector PSP 6K-13-7.  
SQ Sequence 2026 BP; 442 A; 626 C; 532 G; 425 T; 1 Others;  
Query Match 12.7%; Score 7; DB 1; Length 2026;  
Best Local Similarity 100.0%; Pred. No. 7.8e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 49 CTGCACG 55  
DB 1722 CTGCACG 1728  
RESULT 4  
003206/c  
ID Q03206 standard; DNA: 117 BP.

AC 003206;  
DT 10-JUL-1990 (first entry)  
DE Gene encoding protein associated with ras oncogene.  
KW ras oncogene; cancer; GTPase; ds.  
OS Synthetic.  
PN MO9000179-A.  
PD 11-JAN-1990.  
PF 4-JUL-1989; F00348.  
PR (INRW) Inserm Inst Nat Sante.  
PI Tavitian A, Pizon V, Chardin P;  
DR WPI: 90-037122/05.  
DR P-PSDB: R05070.  
PT DNA encoding sequences associated with human oncogenes -  
PT and derived antibodies, useful for in vitro diagnosis of cancer  
PT Claim 8; Page 59; 92pp; French.  
PS Claim 8; Page 59; 92pp; French.  
CC Derived peptides can fix GTP and GDP and have GTPase activity. Abs raised  
CC to these peptides can diagnose associated diseases - absence or excess of  
CC the product indicates loss of balance between anti- and oncogenic factors  
CC in the cells. The product may be used to treat conditions associated with  
CC ras or rap genes.  
SQ Sequence 117 BP; 24 A; 36 C; 34 G; 23 T;  
Query Match 12.7%; Score 7; DB 1; Length 117;  
Best Local Similarity 100.0%; Pred. No. 8.1e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 49 CTGCACG 55  
DB 15 CTGCACG 9

RESULT 5  
003212/c  
ID Q03212 standard; DNA: 558 BP.  
AC Q03212;  
DT 10-JUL-1990 (first entry)  
DE RAP2 Gene encoding at least one peptide associated with ras oncogene.  
KW ras oncogene; cancer; GTPase; ds.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT cds 4.552  
/\*tag- a  
PN MO9000179-A.  
PD 11-JAN-1990.  
PF 4-JUL-1989; F00348.  
PR 4-JUL-1988; FR-009031.  
PA (INRW) Inserm Inst Nat Sante.  
PI Tavitian A, Pizon V, Chardin P;  
DR WPI: 90-037122/05.  
DR P-PSDB: R05076.  
PT DNA encoding sequences associated with human oncogenes -  
PT and derived antibodies, useful for in vitro diagnosis of cancer  
PT and for therapy.  
PS Disclosure: Fig 3a-b; 92pp; French.  
CC Derived peptides can fix GTP and GDP and have GTPase activity. Abs raised  
CC to these peptides can diagnose associated diseases - absence or excess of  
CC the product indicates loss of balance between anti- and oncogenic factors  
CC in the cells. The product may be used to treat conditions associated with  
CC ras or rap genes.  
SQ Sequence 558 BP; 142 A; 143 C; 165 G; 108 T;  
Query Match 12.7%; Score 7; DB 1; Length 558;  
Best Local Similarity 100.0%; Pred. No. 8e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 49 CTGCACG 55  
DB 69 CTGCACG 63

## RESULT 6

003247/c  
ID 003247 standard; DNA; 1766 BP.  
AC 003247;  
DT 11-JUL-1990 (first entry)  
DE Aspergillopepsin A gene.  
KW Aspartic proteinase; chymosin; t-PA; human serum albumin ds.  
OS Aspergillus awamori.  
FH Key Location/Qualifiers  
FT signal\_peptide 178..489 /tag- a  
FT precursor\_rna 238..385 /tag- b  
FT /product-Highly charged propeptide.  
FT cds 178..1521 /tag- c  
FT mat\_peptide 386..1521 /tag- d  
FT intron 490..540 /tag- e  
FT intron 827..878 /tag- f  
FT intron 1128..1185 /tag- g  
FT poly\_a\_site 1573..1577 /tag- h  
FT /evidence-Putative.  
PN W09000192-A.  
PD 11-JAN-1990.  
PF 01-JUL-1989; U02891.  
PR 01-JUL-1988; US-214237.  
PA (GENE-) Genencor Inc.  
PI Berka R, Hayenga K, Lawlis VB, Ward M;  
DR WPI: 90-037128/05.  
DR P-PSDB: R05078.  
PT Aspartic proteinase deficient filamentous fungi prepn. -  
PT by pref. using a gene replacement vector, useful as hosts for  
PT increased prodn. of heterologous polypeptide(s).  
PS Disclosure; Fig 3; 64pp; English.  
CC Mutating the aspergillopepsin gene, a strain deficient in the aspartic  
CC proteinase may be synthesised. The absence of such proteolytic enzymes  
CC improves the production of heterologous gene products, for instance  
CC bovine chymosin, t-PA; human serum albumin or mammalian growth hormone  
CC Sequence 1766 BP; 388 A; 529 C; 430 G; 419 T;

Query Match 12.7%; Score 7; DB 1; Length 1766;  
Best Local Similarity 100.0%; Pred. No. 7.8e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTCACG 55  
DB 951 CTCACG 945

RESULT 7  
ID 003216 standard; DNA; 2122 BP.  
AC 003216;  
DT 11-JUL-1990 (first entry)  
DE SalI restriction fragment of plasmid pUC-RNSII containing rabies  
DE preglycoprotein G protein.  
KM Rabies; pUC-RNSII; pre-glycoprotein G; ds.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT cds 63..1635 /tag- a  
PN W09000191-A.  
PD 11-JAN-1990.  
PF 27-JUN-1989; U02809.  
PR 01-JUL-1988; US-214085.

PA (SALK) Salk Inst for Biol Stud.  
PI Tschopp JF, Dixon DB;  
DR WPI: 90-037127/05.  
DR P-PSDB: R05078.  
PT Prodn. of rabies glyco-protein G -  
PT by culturing P.pastoris contg. DNA encoding preprotein with  
PT methanol as carbon source.  
PS Disclosure; Table 1; 82pp; English.  
CC Isolating rabies preglycoprotein G, and associating it with a promoter  
CC segment of a methyloctrophic yeast strain and a terminator from a second  
CC methyloctrophic strain, allows the protein to be selectively expressed in  
CC P. pastoris culture (pref. ARCC 20880).  
CC The inactive virus protein is useful in a vaccine as a prophylactic  
CC measure for at risk individuals and animals.  
SQ Sequence 2122 BP; 634 A; 458 C; 487 G; 543 T;

Query Match 12.7%; Score 7; DB 1; Length 2122;  
Best Local Similarity 100.0%; Pred. No. 7.7e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTCACG 55  
DB 899 CTCACG 905

RESULT 8  
ID 003246/c  
AC 003246;  
DT 12-JUL-1990 (first entry)  
DE PelC Pectin lyase PLC recombinant DNA molecule.  
KM Pectin lyase; PLC; expression system; interferon.  
KW ds.  
OS Aspergillus niger.  
FH Key Location/Qualifiers  
FT cds 1368..2707.  
FT intron /tag- a  
FT intron 1854..1931 /tag- b  
FT intron 1998..2062 /tag- c  
FT intron 2234..2296 /tag- d  
PN EP-353188-A.  
PD 31-JAN-1990.  
PF 19-JUL-1989.  
PR 28-JUL-1988; GB-018046.  
PR 26-JUN-1989; GB-014666.  
PA (CIBA) Ciba Geigy AG.  
PI Helm J, Meyhack B, Vissler J;  
DR WPI: 90-031714/05.  
DR P-PSDB: R05112.  
PT DNA coding for pectin lyase(s) PLA, PLB, PLC, PLF or PLF -  
PT used to produce pectin lyase(s) in pure form for constructing  
PT hybrid vectors expressing foreign genes.  
PS Claim 5; Fig 12; 63pp; English.  
CC Useful in production of hybrid vectors expressing the pectin lyase  
CC gene and/or hybrid vectors expressing foreign genes eg. interferon  
CC within a filamentous fungal expression system esp. Aspergillus niger.  
SQ Sequence 3168 BP; 823 A; 780 C; 720 G; 845 T;

Query Match 12.7%; Score 7; DB 1; Length 3168;  
Best Local Similarity 100.0%; Pred. No. 7.6e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTCACG 55  
DB 1823 CTCACG 1817

RESULT 9



```

FT      /*tag= a
FT      /label-pre-TGF beta 1
FT      1677..2011
FT      cds
FT      /*tag= b
FT      /label-mature TGF-beta 1
FT      2015..2092
FT      gC-signal
FT      /*tag= c
FT      2093..2099
FT      misc-feature
FT      /*tag= d
FT      /label-TARA-like sequence
FT      stem_loop
FT      37..113
FT      /*tag= e
FT      863..911
FT      misc-feature
FT      /*tag= f
FT      /label-hydrophobic domain
PN      W08912101.A.
PD      14-DEC-1989.
PD      08-JUN-1988; WO-01945.
PD      08-JUN-1988; WO-01945.
PI      (GETH) Genentech Inc.
PI      Dernyck RM, Goeddel DV;
PI      WPI: 90-007474/01.
PI      P-PSDB: R04034.
PT      Nucleotide sequence encoding transforming growth factor beta-3 -used as a
PT      probe, or to produce TGF beta 3, for growth inhibition of certain normal
PT      and neoplastic cells, eg A549.
PS      Disclosure; Fig. 1b; 61pp; English.
CC      Sequence encodes the 390 amino acid (AA) precursor transforming growth
CC      factor-beta 1 (pre-TGF-beta 1) polypeptide. The 5' untranslated region of
CC      the TGF-beta 1 mRNA is 841 bases long, is purine rich and has a region of
CC      potential secondary structure. The TARA-like sequence in the 3' untrans-
CC      lated region of the gene is presumably a polyadenylation signal. Mature
CC      TGF-beta 1 comprises the C-terminal 112 AA's of pre-TGF-beta 1 and is
CC      cleaved at the Arg-Arg dipeptide preceding its NH2 terminus. The nucleic
CC      acid encoding the second subtype of TGF-beta (TGF-beta 3) is useful as a
CC      probe or to produce TGF-beta 3 for inhibition of growth of normal and
CC      neoplastic cells.
SQ      Sequence 2537 BP; 473 A; 893 C; 739 G; 432 T;

Query Match
Best Local Similarity 12.7%; Score 7; DB 1; Length 2537;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      49 CTGCACG 55
        |||||||
        493 CTGCACG 499

RESULT 13
ID      002819 standard; DNA; 2676 BP.
AC      002819;
DT      10-MAR-1993 (revised)
DT      31-MAY-1989 (first entry)
DE      cDNA sequence encoding porcine TGF-beta 3.
DE      Transforming growth factor beta-3 (TGF beta 3); tumour cells; growth
DE      inhibition.
KW      Inhibition.
PN      W08912101.A.
PD      14-DEC-1989.
PD      08-JUN-1988; WO-01945.
PD      08-JUN-1988; WO-01945.
PI      (GETH) Genentech Inc.
PI      Dernyck RM, Goeddel DV;
PI      WPI: 90-007474/01.
PI      P-PSDB: R04080.
PT      Nucleotide sequence encoding transforming growth factor beta-3 -used as a
PT      probe, or to produce TGF beta 3, for inhibition of growth of normal
PT      and neoplastic cells, eg A549.
PS      Disclosure; Fig. 4; 61pp; English.
CC      This sequence encodes porcine transforming growth factor-beta 3 (TGF-
CC      beta 3) polypeptide. The nucleic acid sequence encoding this subtype is
CC      useful as a probe or to produce TGF-beta 3 for both normal and neoplastic

```

```

CC      cell growth inhibition.
SQ      Sequence 2676 BP; 704 A; 705 C; 699 G; 568 T;

Query Match
Best Local Similarity 12.7%; Score 7; DB 1; Length 2676;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      49 CTGCACG 55
        |||||||
        193 CTGCACG 199

Db      193 CTGCACG 199

RESULT 14
ID      003218 standard; DNA; 1389 BP.
AC      003218;
DT      10-MAR-1993 (revised)
DT      12-JUL-1989 (first entry)
DE      DNA sequence encoding human protein C mutant
DE      Human protein C; anti-coagulant; ss.
DE      Homo sapiens.
KW      Key
KW      Location/Qualifiers
FH      Key
FT      cds
FT      1..1383
FT      /*tag= a
FT      /label-Human protein C mutant
PN      EP-352651-A.
PD      31-JAN-1990.
PD      21-JUL-1989; 113414.
PD      26-JUL-1988; JP-184538.
PI      (FARU) Hoechst Japan Ltd.
PI      Hashimoto T, Sato M;
PI      WPI: 90-031241/05.
PI      P-PSDB: R04800.
PT      Human protein C mutant with residues 156 and 157 replaced or deleted -
PT      has reduced activation rate and extended duration of activity as
PT      anti-coagulant.
PS      Disclosure; Page 11-14; 22pp; English.
CC      This sequence encodes a protein which differs from human protein C
CC      at positions 156 and 157 (Lys-Arg is substituted by Asn-Ser). This
CC      creates a single uncleaved chain. The mutant product shows anti-coagulant
CC      activity when activated by the thrombin-thrombomodulin complex. It has
CC      a lower activation rate and longer duration of activity compared with
CC      the native protein.
SQ      Sequence 1389 BP; 289 A; 419 C; 433 G; 248 T;

Query Match
Best Local Similarity 12.7%; Score 7; DB 1; Length 1389;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      49 CTGCACG 55
        |||||||
        438 CTGCACG 444

Db      438 CTGCACG 444

RESULT 15
ID      001794 standard; DNA; 1383 BP.
AC      001794;
DT      25-JUL-1990 (first entry)
DE      Nucleotide sequence coding for hybrid human/bovine protein C
DE      Human protein C; bovine protein C; Gla domain; blood anticoagulant;
DE      blood coagulation disorders; blood fibrinolysis acceleration.
KW      EP-354504-A.
PN      14-FEB-1990.
PD      05-AUG-1988; 114505.
PD      09-AUG-1988; JP-197144.
PI      (FARU) Hoechst Japan Ltd.
PI      Hashimoto T, Takahashi M;
PI      WPI: 90-046218/07.
PI      P-PSDB: R05211.
PT      Human protein C deriv. -

```



PT having Gla domain replaced by the bovine Gla to increase calcium  
 binding activity and improve activation  
 PS Claim 3, Page 14, 17pp; English.  
 CC Also claimed are its encoded human protein C, the gene structure contg.  
 CC it and a host cell contg. the gene structure. The N-terminal region of  
 CC human protein C having gamma-carboxylated glutamic acid residues (Gla  
 CC domain) has been replaced by the Gla domain of bovine protein C or by an  
 CC equivalent of this bovine protein sequence w.r.t. its calcium binding  
 CC activity, and/or its enhanced protein C activity. PCS4 was constructed  
 CC from PCS1. DNA coding for part of protein C was removed from PCS4 by  
 CC Sall digestion at a site just upstream of the leader sequence-coding  
 CC sequence and again at a site corresp. to the 45th (Val) and 46th (Asp)  
 CC residues of protein C. A synthetic DNA fragment was inserted into PCS4  
 CC between these Sall sites and introduced into E. coli K12/HR101. The  
 CC transformants harbouring a plasmid in an appropriate orientation, named  
 CC PCS8, were screened and cultivated. PHSG293 was also constructed which  
 CC had a neo gene as a selective marker and BstXI site providing the same  
 CC asymmetric cohesive ends as PCS8 by digestion. The PCS8 was digested  
 CC with BstXI and the 2 kb fragment was ligated with BstXI-digested PHSG293.  
 CC The ligated DNA was packaged in vitro in a lambda phage packaging mixt.  
 CC and transfected into E. coli K12/Om206. The DNA was isolated and  
 CC introduced into CHO cells which were cultured to produce the hybrid  
 CC protein C. Activated protein C inhibits blood coagulation or accelerates  
 CC blood fibrinolysis and is used for treating blood coagulation disorders.  
 SQ Sequence 1383 BP; 288 A; 413 C; 442 G; 240 T;

Query Match 12.78; Score 7; DB 1; Length 1383;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+03;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTGCACG 55  
 DB 432 CTGCACG 438

Search completed: September 18, 1999, 00:33:26  
 Job time: 18966 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:27:52 : Search time 2825.05 Seconds  
(without alignments)  
6.982 Million cell updates/sec

Title: US-08-956-518a-85  
Perfect score: 10  
Sequence: 1 TCRCCTTAG 10

Scoring table: IDENTITY\_NUC

arched: 2546578 seqs, 98626752 residues

-abase :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: em\_est20: \*  
21: em\_est21: \*  
22: em\_est22: \*  
23: em\_est23: \*  
24: em\_est24: \*  
25: em\_est25: \*  
26: em\_est26: \*  
27: em\_est27: \*  
28: em\_est28: \*  
29: em\_est29: \*  
30: em\_est30: \*  
31: em\_est31: \*  
32: em\_est32: \*  
33: em\_est33: \*  
34: em\_est34: \*  
35: em\_est35: \*  
36: em\_est36: \*  
37: em\_est37: \*  
38: em\_est38: \*  
39: em\_est39: \*  
40: em\_est40: \*  
41: em\_est41: \*  
42: em\_est42: \*  
43: em\_est43: \*  
44: em\_est44: \*  
45: em\_est45: \*  
46: em\_est46: \*  
47: em\_est47: \*  
48: em\_est48: \*  
49: em\_est49: \*  
50: em\_est50: \*  
51: em\_est51: \*  
52: em\_est52: \*  
53: em\_est53: \*

54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	100.0	38	20	D15131	D15131	RICCO129A R
2	100.0	435	20	D15980	D15980	RICCO1808A R
3	100.0	405	20	D21672	D21672	MUS81D09 MO
4	100.0	384	20	D22924	D22924	RICCI194A R
5	100.0	347	20	D24072	D24072	RICRI057A R
6	100.0	430	20	D24444	D24444	RICRI032A R
7	100.0	318	20	D24993	D24993	RICR2912A R
8	100.0	205	20	D25754	D25754	HUMGS04121
9	100.0	310	20	D31273	D31273	HUMI13202 H
10	100.0	414	20	D34553	D34553	CELR0082R
11	100.0	315	20	D35356	D35356	CELR001DYF
12	100.0	360	20	D35439	D35439	CELR020E5F
13	100.0	330	20	D36191	D36191	CELR002GZF
14	100.0	444	20	D38872	D38872	RICCI180A R
15	100.0	407	20	D39092	D39092	RICRI16A R
16	100.0	344	20	L26626	L26626	MUSB045A 1a
17	100.0	252	20	L36676	L36676	BNAESTF756
18	100.0	376	20	M62140	M62140	ESTP00201 H1
19	100.0	163	20	M62160	M62160	ESTP00222 H1
20	100.0	448	20	M62264	M62264	ESTP00333 H1
21	100.0	410	20	M79639	M79639	WEST00176 M
22	100.0	460	20	M79767	M79767	WEST00304 M
23	100.0	417	20	M79768	M79768	WEST00305 M
24	100.0	422	20	M79840	M79840	WEST00377 M
25	100.0	427	20	T00381	T00381	WEST01102 E
26	100.0	339	20	T00955	T00955	WEST01680 E
27	100.0	406	20	T00997	T00997	WEST01718 E
28	100.0	176	20	T01374	T01374	WEST02095 E
29	100.0	357	20	T01649	T01649	WEST02370 E
30	100.0	417	20	T02220	T02220	WEST02941 E
31	100.0	311	20	T05175	T05175	ESTP03063 Fe
32	100.0	351	20	T06527	T06527	ESTP04416 Fe
33	100.0	398	20	T06573	T06573	ESTP04462 Fe
34	100.0	374	20	T06602	T06602	ESTP04491 Fe
35	100.0	354	20	T08281	T08281	ESTP06172 In
36	100.0	393	20	T09065	T09065	ESTP06958 In
37	100.0	194	20	T10525	T10525	hbc706 Huma
38	100.0	372	20	T13618	T13618	1783 Lambda
39	100.0	235	20	T16466	T16466	NIB1349 Nor
40	100.0	330	20	T18018	T18018	0360C3 CBSP
41	100.0	579	20	T18334	T18334	GUP 8-3 all
42	100.0	202	20	T20053	T20053	B360F Heart
43	100.0	349	20	T24010	T24010	seg2166 3HF
44	100.0	455	20	T24372	T24372	CRS1502 1am
45	100.0	317	51	AT15497	AT15497	UI-R-YO-a

#### ALIGNMENTS

RESULT	1	D15131	328 bp	EST	20-JUL-1998
LOCUS		D15131			
DEFINITION		RICCO129A	Rice callus	Oriza sativa	CDNA clone R, mRNA sequence.
ACCESSION		D15131			
NID		5286323			
VERSION		D15131.1			GI:286323
KEYWORDS		EST.			

SOURCE  
ORGANISM  
Oryza sativa.  
Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
Poaceae; Oryza.  
REFERENCE  
1 (bases 1 to 328)  
Sasaki, T. and Minobe, Y.  
Rice cDNA from callus  
Unpublished (1994)  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Contact: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program  
2-1-2 Kannondai, Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441  
Fax: 0298-38-7468  
Email: tsasaki@abr.affrc.go.jp  
PROJECT = "RGP"  
Seq primer: oligo(dT).  
Location/Qualifiers  
1..328  
/organism="Oryza sativa"  
/strain="cultivar Nipponbare, sub-species Japonica"  
/db\_xref="taxon:4530"  
/clone="R"  
/clone\_lib="Rice callus"  
/note="Vector: pBluescript II SK+; Site\_1: SalI; Site\_2:  
NotI; cDNA prepared from rice callus mRNAs by using  
oligo(dT) as a primer and ligating to the SalI-NotI site  
of pBluescript II SK+ phagemid.  
3 others  
BASE COUNT  
73 a 100 c 101 g 51 t  
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 328;  
Best Local Similarity 100.0%; Pred. No. 4.3e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TCTCCTTAG 10  
|||||  
Db 222 TCTCCTTAG 213

RESULT 2  
D15980 435 bp mRNA EST 20-JUL-1998  
LOCUS  
INITIATION RICC1808A Rice callus Oryza sativa cDNA clone R. mRNA sequence.  
D15980  
NID 9287175  
VERSION D15980.1 GI:287175  
KEYWORDS  
SOURCE EST.  
ORGANISM Oryza sativa.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
Poaceae; Oryza.  
REFERENCE  
1 (bases 1 to 435)  
Sasaki, T. and Minobe, Y.  
Rice cDNA from callus  
Unpublished (1994)  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Contact: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program  
2-1-2 Kannondai, Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441  
Fax: 0298-38-7468  
Email: tsasaki@abr.affrc.go.jp  
PROJECT = "RGP"

Seq primer: oligo(dT).  
Location/Qualifiers  
1..435  
/organism="Oryza sativa"  
/strain="cultivar Nipponbare, sub-species Japonica"  
/db\_xref="taxon:4530"  
/clone="R"  
/clone\_lib="Rice callus"  
/note="Vector: pBluescript II SK+; Site\_1: SalI; Site\_2:  
NotI; cDNA prepared from rice callus mRNAs by using  
oligo(dT) as a primer and ligating to the SalI-NotI site  
of pBluescript II SK+ phagemid.  
1 others  
BASE COUNT  
108 a 97 c 117 g 112 t  
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 435;  
Best Local Similarity 100.0%; Pred. No. 4.4e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TCTCCTTAG 10  
|||||  
Db 70 TCTCCTTAG 79

RESULT 3  
D21672 405 bp mRNA EST 07-OCT-1996  
LOCUS  
DEFINITION MUS81D09 mouse embryonal carcinoma cell line F9 Mus musculus cDNA  
clone 81D09, mRNA sequence.  
D21672  
ACCESSION 9618800  
NID D21672  
VERSION D21672.1 GI:618800  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 405)  
Nishiguchi, S., Joh, T., Horie, K., Zou, Z., Yasunaga, T. and Shimada, K.  
A survey of genes expressed in undifferentiated mouse embryonal  
carcinoma F9 cells: characterization of low-abundance mRNAs  
J. Biochem. 116, 128-139 (1994)  
95096008  
JOURNAL  
MEDLINE  
COMMENT  
Contact: Kazunori Shimada  
Department of Medical Genetics, Division of Molecular Biomedicine  
Research Institute for Microbial Diseases, Osaka University  
3-1, Yamadaoka, Suita, Osaka, 565, Japan  
Tel: 06-879-8325  
Fax: 06-879-8326.  
Location/Qualifiers  
1..405  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="81D09"  
/clone\_lib="mouse embryonal carcinoma cell line F9"  
BASE COUNT  
130 a 74 c 117 g 84 t  
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 405;  
Best Local Similarity 100.0%; Pred. No. 4.4e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TCTCCTTAG 10  
|||||  
Db 157 TCTCCTTAG 148

RESULT 4  
D22924 394 bp mRNA EST 20-JUL-1998  
LOCUS

```

DEFINITION  RICCI1794A Rice callus Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION   D22924
NID         9426849
VERSION     D22924.1 GI:426849
KEYWORDS    EST.
SOURCE      Oryza sativa.
ORGANISM    Oryza sativa.
REFERENCE   1 (bases 1 to 394)
AUTHORS    Sasaki,T. and Minobe,Y.
TITLE      Rice cDNA from callus
JOURNAL    Unpublished (1994)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT = 'RGP'

FEATURES
  source
    1..394
    /organism="Oryza sativa"
    /strain="Cultivar Nipponbare, sub_species Japonica"
    /db_xref="taxon:4530"
    /clone="R"
    /clone_1lb="Rice callus"
    /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
    NotI; cDNA prepared from rice callus mRNAs by using
    oligo(dT) as a primer and ligating to the SalI-NotI site
    of pBluescript II SK+ phagemid. 4 others
    122 c 117 g 61 t

BASE COUNT  90 a 122 c 117 g 61 t 4 others
ORIGIN

Query Match      100.0%; Score 10; DB 20; Length 394;
Best Local Similarity 100.0%; Pred. No. 4.3e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAG 10
    |||||
    221 TCTCCTTAG 212

RESULT 5
D24072/c 347 bp mRNA EST 20-JUL-1998
LOCUS    RICRI057A Rice root Oryza sativa cDNA clone R, mRNA sequence.
DEFINITION
D24072
NID      9427928
VERSION  D24072.1 GI:427928
KEYWORDS EST.
SOURCE   Oryza sativa.
ORGANISM Oryza sativa.
REFERENCE 1 (bases 1 to 347)
AUTHORS  Minobe,Y. and Sasaki,T.
TITLE    Rice cDNA from root
JOURNAL  Unpublished (1995)
COMMENT  Contact: Takuji Sasaki
        National Institute of Agrobiological Resources
        Rice Genome Research Program
        2-1-2 Kannondai,Tsukuba
        Ibaraki,
        Japan 305

FEATURES
  source
    1..430
    /organism="Oryza sativa"
    /strain="Nipponbare, sub_species Japonica"
    /db_xref="taxon:4530"
    /clone="R"
    /clone_1lb="Rice root"
    /note="Prepared from seedling root."
    129 a 80 c 106 g 113 t 2 others

BASE COUNT  129 a 80 c 106 g 113 t 2 others
ORIGIN

Query Match      100.0%; Score 10; DB 20; Length 430;
Best Local Similarity 100.0%; Pred. No. 4.4e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAG 10
    |||||
    Db 108 TCTCCTTAG 99

DEFINITION  RICRI932A Rice root Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION   D24444
NID         9428296
VERSION     D24444.1 GI:428296
KEYWORDS    EST.
SOURCE      Oryza sativa.
ORGANISM    Oryza sativa.
REFERENCE   1 (bases 1 to 430)
AUTHORS    Minobe,Y. and Sasaki,T.
TITLE      Rice cDNA from root
JOURNAL    Unpublished (1995)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT = 'RGP'

FEATURES
  source
    1..430
    /organism="Oryza sativa"
    /strain="Nipponbare, sub_species Japonica"
    /db_xref="taxon:4530"
    /clone="R"
    /clone_1lb="Rice root"
    /note="Prepared from seedling root."
    102 a 60 c 90 g 95 t

BASE COUNT  102 a 60 c 90 g 95 t
ORIGIN

Query Match      100.0%; Score 10; DB 20; Length 347;
Best Local Similarity 100.0%; Pred. No. 4.3e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAG 10
    |||||
    Db 74 TCTCCTTAG 65

RESULT 6
D24444/c 430 bp mRNA EST 20-JUL-1998
LOCUS    RICRI932A Rice root Oryza sativa cDNA clone R, mRNA sequence.
DEFINITION
D24444
NID      9428296
VERSION  D24444.1 GI:428296
KEYWORDS EST.
SOURCE   Oryza sativa.
ORGANISM Oryza sativa.
REFERENCE 1 (bases 1 to 430)
AUTHORS  Minobe,Y. and Sasaki,T.
TITLE    Rice cDNA from root
JOURNAL  Unpublished (1995)
COMMENT  Contact: Takuji Sasaki
        National Institute of Agrobiological Resources
        Rice Genome Research Program
        2-1-2 Kannondai,Tsukuba
        Ibaraki,
        Japan 305
        Tel: 0298-38-7441
        Fax: 0298-38-7468
        Email: tsasaki@abr.affrc.go.jp
        PROJECT = 'RGP'

FEATURES
  source
    1..430
    /organism="Oryza sativa"
    /strain="Nipponbare, sub_species Japonica"
    /db_xref="taxon:4530"
    /clone="R"
    /clone_1lb="Rice root"
    /note="Prepared from seedling root."
    102 a 60 c 90 g 95 t

BASE COUNT  102 a 60 c 90 g 95 t
ORIGIN

```

RESULT 7  
D24993/c 318 bp mRNA EST 20-JUL-1998  
LOCUS R1CR2912A Rice root Oryza sativa cDNA clone R. mRNA sequence.  
DEFINITION D24993  
ACCESSION 9428837  
NID D24993.1 GI:428837  
VERSION EST.  
KEYWORDS Oryza sativa.  
SOURCE Oryza sativa.  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.  
REFERENCE 1 (bases 1 to 318)  
AUTHORS Minobe,Y. and Sasaki,T.  
TITLE Rice cDNA from root.  
JOURNAL Unpublished (1995)  
COMMENT Contact: Yuzo Minobe  
National Institute of Agrobiological Resources  
Rice Genome Research Program  
2-1-2 Kannondai,Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441  
Fax: 0298-38-7468  
Email: minobe@rcs0.riken.go.jp  
PROJECT "RGP"

FEATURES  
Source  
1..318  
/organism="Oryza sativa"  
/strain="Nipponbare, sub-species Japonica"  
/db\_xref="taxon:4530"  
/clone="R"  
/clone\_1lb="Rice root"  
/note="Prepared from seedling root."

BASE COUNT 96 a 54 c 81 g 85 t 2 others  
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 318;  
Best Local Similarity 100.0%; Pred. No. 4.3e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAG 10  
|||||  
108 TCTCCTTAG 99

RESULT 8  
D25754/c 205 bp mRNA EST 30-NOV-1995  
LOCUS HUMS04121 Human colon mucosa Homo sapiens cDNA clone cml888 3',  
DEFINITION D25754  
ACCESSION 9500443  
NID D25754.1 GI:500443  
VERSION EST.  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 205)  
AUTHORS Okubo,K., Yoshii,Y., Yokouchi,H., Kameyama,M. and Matsubara,K.  
TITLE Global analysis of gene expression in colon mucosa: a large scale random cDNA sequencing analysis  
JOURNAL Unpublished (1994)  
COMMENT Contact: Okubo,K., Itoh,K., Yoshii,Y., Yokouchi,H. and Matsubara,K.  
Institute for Molecular and Cellular Biology  
Osaka University  
3-1 Yamada-oka,Suita,Osaka 565,Japan.

FEATURES  
source  
1..205  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="cml888"  
/clone\_1lb="Human colon mucosa"  
/note="Adult male, tissue type = colon mucosa"

BASE COUNT 53 a 44 c 51 g 52 t 5 others  
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 205;  
Best Local Similarity 100.0%; Pred. No. 4.2e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAG 10  
|||||  
Db 291 TCTCCTTAG 300

RESULT 10  
D34553 414 bp mRNA EST 05-AUG-1994  
LOCUS CELK008B2R Yuj1 Kohara unpublished cDNA Caenorhabditis elegans cDNA  
DEFINITION D34553  
ACCESSION 9522797  
NID

FEATURES  
source  
1..310  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1lb="Human fetal lung"

BASE COUNT 67 a 86 c 78 g 78 t 1 others  
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 310;  
Best Local Similarity 100.0%; Pred. No. 4.3e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAG 10  
|||||  
Db 291 TCTCCTTAG 300

RESULT 10  
D34553 414 bp mRNA EST 05-AUG-1994  
LOCUS CELK008B2R Yuj1 Kohara unpublished cDNA Caenorhabditis elegans cDNA  
DEFINITION D34553  
ACCESSION 9522797  
NID

COMMENT Contact: Yusuke Nakamura  
Institute of Medical Science  
University of Tokyo  
4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan  
Tel: 81-3-5449-5372  
Fax: 81-3-5449-5433  
Email: yusuke@ims.u-tokyo.ac.jp  
Insert length: 596 Std Error: 0.00  
High quality sequence stop: 157.  
Location/Qualifiers  
1..310  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1lb="Human fetal lung"

```

VERSION      D34553.1  GI:522797
KEYWORDS     EST
SOURCE       D35356/c
ORGANISM     Caenorhabditis elegans.
REFERENCE    1 (bases 1 to 414)
AUTHORS      Kohara,Y., Mitsukl,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
              Tabara,H.
TITLE        Toward an expression map of the C.elegans genome
JOURNAL      Unpublished (1994)
COMMENT

FEATURES
  source      Contact: Yuji Kohara
              Gene Library Lab
              National Institute of Genetics
              Yata 1111, Mishima, Shizuoka 411, Japan
              Tel: 0559-75-0771
              Fax: 0559-75-6240
              Email: ykohara@dbj.nig.ac.j.
              Location/Qualifiers
                1..414
                /organism="Caenorhabditis elegans"
                /strain="CB1489 him-8(e1489)"
                /db_xref="taxon:6239"
                /clone="YK8b2"
                /clone_lib="Yuji Kohara unpublished cDNA"
                /note="dev_stage=varied, sex=Hermaphrodite male,
                tissue_type=whole animal"
                99 a 92 c 109 g 112 t 2 others

BASE COUNT   99 a 92 c 109 g 112 t 2 others
ORIGIN

Query Match      100.0%; Score 10; DB 20; Length 414;
Best Local Similarity 100.0%; Pred. No. 4.4e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

Oy      1 TCTCCTAAG 10
        |||||
Db      301 TCTCCTAAG 310

RESULT 11
D35356/c
LOCUS      D35356      315 bp      mRNA      EST      08-AUG-1994
DEFINITION CEK001DIF Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
LOCUS      D35356      clone yk1d11 5', mRNA sequence.
SSION      9525357
          D35356.1  GI:525357
VERSION    EST.
SOURCE     Caenorhabditis elegans.
ORGANISM   Caenorhabditis elegans.
KEYWORDS   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
          Rhabditina; Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE  1 (bases 1 to 315)
AUTHORS    Kohara,Y., Mitsukl,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
          Tabara,H.
TITLE      Toward an expression map of the C.elegans genome
JOURNAL    Unpublished (1994)
COMMENT

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.j.
Location/Qualifiers
  1..315
  /organism="Caenorhabditis elegans"
  /strain="CB1489 him-8(e1489)"
  /db_xref="taxon:6239"
  /clone="yk1d11"

```

BASE COUNT	82 a	101 c	68 g	62 t	2 others
ORIGIN					
Query Match	100.0%;	Score 10;	DB 20;	Length 315;	
Best Local Similarity	100.0%;	Pred. No. 4.3e+03;			
Matches 10;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
OY	1	TCTCCTTAAG 10			
Db	299	TCTCCTTAAG 290			
RESULT 12					
D35439/c	360 bp	mRNA	EST	08-AUG-1994	
LOCUS	CEIK0205EF	Yuj1 Kohara unpublished cDNA	Caenorhabditis elegans	cdNA	
DEFINITION	clone yK20e5 5',	mRNA sequence.			
ACCESSION	D35439				
NID	9526890				
VERSION	D35439.1	GI:526890			
KEYWORDS	EST.				
SOURCE	Caenorhabditis elegans.				
ORGANISM	Caenorhabditis elegans				
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidae; Rhabditidae; Peloderiinae; Caenorhabditis.				
AUTHORS	1 (bases 1 to 360)				
	Kohara,Y., Mitsukl,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and Nabara,H.				
TITLE	Toward an expression map of the C.elegans genome				
JOURNAL	Unpublished (1994)				
COMMENT	Contact: Yuj1 Kohara				
	Gene Library Lab				
	National Institute of Genetics				
	Yata 111, Mishima, Shizuoka 411, Japan.				
	Tel: 0559-75-0771				
	Fax: 0559-75-6240				
	Email: ykoha@dbj.nig.ac.jp.				
FEATURES	Location/Qualifiers				
Source	1..360				
	/organism="Caenorhabditis elegans"				
	/strain="CB1489 him-8(e1489)"				
	/db_xref="taxon:6239"				
	/clone="yK20e5"				
	/clone_id="Yuj1 Kohara unpublished cDNA"				
	/note="dev_stage=varied, sex=hermaphrodite male, tissue-type=whole animal"				
BASE COUNT	111 a	77 c	79 g	90 t	3 others
ORIGIN					
Query Match	100.0%;	Score 10;	DB 20;	Length 360;	
Best Local Similarity	100.0%;	Pred. No. 4.3e+03;			
Matches 10;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
OY	1	TCTCCTTAAG 10			
Db	352	TCTCCTTAAG 343			
RESULT 13					
D36191/c	330 bp	mRNA	EST	08-AUG-1994	
LOCUS	CEIK0026ZF	Yuj1 Kohara unpublished cDNA	Caenorhabditis elegans	cdNA	
DEFINITION	clone yK2g12 5',	mRNA sequence.			
ACCESSION	D36191				
NID	9525409				
VERSION	D36191.1	GI:525409			
KEYWORDS	EST.				

SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Caenorhabditis elegans.  
Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;  
Rhabditina; Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.  
1 (bases 1 to 330)  
Kohara, Y., Mitsuki, H., Nishigaki, A., Motonashi, T., Sugimoto, A. and  
Tabara, H.  
Toward an expression map of the C.elegans genome  
Unpublished (1994)

Contact: Yuji Kohara  
Gene Library Lab  
National Institute of Genetics  
Yata 111, Mishima, Shizuoka 411, Japan  
Tel: 0559-75-0771  
Fax: 0559-75-6240  
Email: ykohara@dbi.nig.ac.jp.  
Location/Qualifiers

FEATURES  
Source  
1. 330  
/organism="Caenorhabditis elegans"  
/strain="CB1489 him-8(e1489)"  
/db\_xref="taxon:6239"  
/clone="yk2g12"  
/clone\_1lb="Yuji Kohara unpublished cDNA"  
/note="dev-stage-varied, sex-Hermaphrodite male,  
tissue-type="whole animal"

BASE COUNT  
ORIGIN

79 a 89 c 75 g 72 t 15 others

Query Match  
Best Local Similarity 100.0%; Score 10; DB 20; Length 330;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCTCCTTAG 10  
|||||  
Db 101 TCTCCTTAG 92

RESULT 14  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

D38872 444 bp mRNA EST 11-NOV-1994  
R1C1810A Rice callus Oryza sativa cDNA, mRNA sequence.  
D38872  
9568037  
D38872.1 GI:568037  
EST.  
Oryza sativa.  
Oryza sativa.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
Poaceae; Oryza.  
1 (bases 1 to 444)  
Sasaki, T., Miyao, A. and Yamamoto, K.  
Rice cDNA from callus 1995  
Unpublished (1995)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Contact: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program  
2-1-2 Kannondai, Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441  
Fax: 0298-38-7468  
Email: tsasaki@abr.affrc.go.jp.  
Location/Qualifiers

FEATURES  
Source  
1. 444  
/organism="Oryza sativa"  
/strain="Cultivar Nipponbare, sub-species Japonica"  
/db\_xref="taxon:4530"  
/clone\_1lb="Rice callus"  
/note="Vector: pBluescript II SK+; Site\_1: SalI; Site\_2:

BASE COUNT  
ORIGIN

111 a 96 c 121 g 114 t 2 others

Query Match  
Best Local Similarity 100.0%; Score 10; DB 20; Length 444;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCTCCTTAG 10  
|||||  
Db 70 TCTCCTTAG 79

RESULT 15  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

D39092 407 bp mRNA EST 11-NOV-1994  
R1C1916A Rice root Oryza sativa cDNA, mRNA sequence.  
D39092  
9568255  
D39092.1 GI:568255  
EST.  
Oryza sativa.  
Oryza sativa.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
Poaceae; Oryza.  
1 (bases 1 to 407)  
Sasaki, T., Miyao, A. and Yamamoto, K.  
Rice cDNA from callus 1995  
Unpublished (1995)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Contact: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program  
2-1-2 Kannondai, Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441  
Fax: 0298-38-7468  
Email: tsasaki@abr.affrc.go.jp.  
Location/Qualifiers

FEATURES  
Source  
1. 407  
/organism="Oryza sativa"  
/strain="Nipponbare, sub-species Japonica"  
/db\_xref="taxon:4530"  
/clone\_1lb="Rice root"  
/note="Prepared from seedling root."

BASE COUNT  
ORIGIN

117 a 70 c 103 g 114 t 3 others

Query Match  
Best Local Similarity 100.0%; Score 10; DB 20; Length 407;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCTCCTTAG 10  
|||||  
Db 226 TCTCCTTAG 217

Search completed: September 17, 1999, 21:27:55  
Job time: 14259 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:01:00 ; Search time 1962.17 Seconds  
(without alignments)  
16.208 Million cell updates/sec

Title: US-08-956-518A-87  
Perfect score: 10  
Sequence: 1 TGTGTCAC 10

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

abase :

GenEmbl: \*  
1: gb\_dal: \*  
2: gb\_ba2: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_p11: \*  
8: gb\_p12: \*  
9: gb\_p13: \*  
10: gb\_p14: \*  
11: gb\_p15: \*  
12: gb\_p16: \*  
13: gb\_p17: \*  
14: gb\_p18: \*  
15: gb\_p19: \*  
16: gb\_p20: \*  
17: gb\_p21: \*  
18: gb\_p22: \*  
19: gb\_p23: \*  
20: gb\_p24: \*  
21: gb\_p25: \*  
22: gb\_p26: \*  
23: gb\_p27: \*  
24: gb\_p28: \*  
25: gb\_p29: \*  
26: gb\_p30: \*  
27: gb\_p31: \*  
28: gb\_p32: \*  
29: gb\_p33: \*  
30: gb\_p34: \*  
31: gb\_p35: \*  
32: gb\_p36: \*  
33: gb\_p37: \*  
34: gb\_p38: \*  
35: gb\_p39: \*  
36: gb\_p40: \*  
37: gb\_p41: \*  
38: gb\_p42: \*  
39: gb\_p43: \*  
40: gb\_p44: \*  
41: gb\_p45: \*  
42: gb\_p46: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100.0	5598	1	AAM5961	AJ005961 Acidianus

2	10	100.0	10498	1	AB008831	AB008831 Acinetob
3	10	100.0	3754	1	AB016431	AB016431 Staphyloc
4	10	100.0	10446	1	AE000810	AE000810 Methanob
5	10	100.0	16189	1	AE000845	AE000845 Methanob
6	10	100.0	10072	1	AE000880	AE000880 Methanob
7	10	100.0	30100	1	AF006655	AF006655 Bacillus
8	10	100.0	29950	1	AF015775	AF015775 Bacillus
9	10	100.0	2259	1	ASAR60	X78654 Anabaena sp
10	10	100.0	5198	1	BACWBA	D10388 Bacillus su
11	10	100.0	2497	1	BACWLB	M81324 Bacillus su
12	10	100.0	2277	1	BACWYMEI	L08809 Bacillus su
13	10	100.0	8397	1	BSU55043	U55043 Bacillus su
14	10	100.0	213190	1	BSUB0004	Z99107 Bacillus su
15	10	100.0	208430	1	BSUB0005	Z99108 Bacillus su
16	10	100.0	208430	1	BSUB0008	Z99111 Bacillus su
17	10	100.0	208330	1	BSUB0009	Z99112 Bacillus su
18	10	100.0	208780	1	BSUB0010	Z99114 Bacillus su
19	10	100.0	207730	1	BSUB0011	Z99118 Bacillus su
20	10	100.0	218410	1	BSUB0015	Z99118 Bacillus su
21	10	100.0	212610	1	BSUB0019	Z99122 Bacillus su
22	10	100.0	7589	1	BSYLOGCOD	AJ000975 Bacillus
23	10	100.0	19712	1	D78509	D78509 Bacillus su
24	10	100.0	20341	1	D86418	D86418 Bacillus su
25	10	100.0	17952	1	D90869	D90869 E.coli geno
26	10	100.0	1549	1	ECNUPC	X74825 E.coli nupC
27	10	100.0	4756	1	ECNUPC	M60177 Escherichia
28	10	100.0	493	1	ECNUPC	M17354 E.coli entF
29	10	100.0	1997	1	ECNUPC	J04216 Escherichia
30	10	100.0	136742	1	ECNUPC	D82598 Escherichia
31	10	100.0	2007	1	HPY18884	Y18884 Helicobacte
32	10	100.0	2020	1	HPY18887	Y18887 Helicobacte
33	10	100.0	1993	1	HPY18988	Y18988 Helicobacte
34	10	100.0	1993	1	HPY18989	Y18989 Helicobacte
35	10	100.0	2010	1	HPY18990	Y18990 Helicobacte
36	10	100.0	6962	1	KPY011907	AJ011907 Klebsiell
37	10	100.0	4972	1	MAMAMTRM	X79027 M.ammoniaph
38	10	100.0	37821	1	MICB1770	Z70722 Mycobacteri
39	10	100.0	35445	1	MICB2052	Z98604 Mycobacteri
40	10	100.0	36310	1	MICB4	AL023514 Mycobacte
41	10	100.0	3822	1	MICB4	Z68206 M.leprie DN
42	10	100.0	36063	1	MICB1133CS	U15184 Mycobacteri
43	10	100.0	42106	1	MSGB1133CS	L78811 Mycobacteri
44	10	100.0	38721	1	MSGY219	AD000013 Mycobacte
45	10	100.0	209925	42	AF002223	AF002223 Human gen

## ALIGNMENTS

RESULT 1  
AAM5961 5598 bp DNA BCT 01-AUG-1998  
LOCUS Acidianus ambivalens schABCD operon, nagd gene, ORF2 and partial  
DEFINITION ORF1 and 3.  
ACCESSION AJ005961  
VERSION 93378536  
NID 93378536  
KEYWORDS 2Fe-2S-ferredoxin; 4Fe-4S-ferredoxin; flavoprotein; iron-sulfur  
protein; nagd gene; ORF1; ORF2; ORF3; schA gene; schABCD operon;  
Acidianus ambivalens.  
SOURCE  
ORGANISM Acidianus ambivalens  
Archaea: Crenarchaeota: Sulfolobales: Acidianus.  
REFERENCE  
AUTHORS Kletzin A.  
TITLE Direct Submission  
JOURNAL Submitted (04-MAY-1998) Kletzin A., Institut fuer Mikrobiologie und  
Genetik, Technische Universitaet Darmstadt, Schnittspahnstrasse 10,  
Darmstadt, 64297, GERMANY  
2 (bases 1 to 5598)  
REFERENCE  
AUTHORS Kletzin A.  
TITLE The succinate dehydrogenase from the extremely thermophilic  
archaeon Acidianus ambivalens



```
JOURNAL Unpublished
COMMENT Related sequence: Y09041.
FEATURES Location/Qualifiers
SOURCE 1..5598
        /organism="Acidianus ambivalens"
        /isolate="Iel 10"
        /db_xref="taxon:2283"
        complement(<1..229)
        /note="ORF1"
        /codon_start=1
        /transl_except="(pos:complement(227..229),aa:Met)
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="CA06777.1"
        /db_xref="PID:e1313719"
        /db_xref="PID:g3378537"
        /db_xref="GI:3378537"
        /db_xref="SPTREMBL:O73934"
        /translation="MDESEKLRKYSKLISPSARILLFDIVEGALIAFRGLIEGIL
        FFSFLIYSVSLAIIIGKRIRIATLTLSISA"
        259..531
        /note="ORF2"
        /codon_start=1
        /transl_except="(pos:259..261,aa:Met)
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="CA06778.1"
        /db_xref="PID:e1313720"
        /db_xref="PID:g3378538"
        /db_xref="GI:3378538"
        /db_xref="SPTREMBL:O73935"
        /translation="MIDREKLGYNKLVNHTLSDQDLVIEDKSGKTYLTFVNNYFKI
        MPAYPGKWDEESLYRPFGLGFVFEDEGINETIKKLEELKSGAL"
        518..1303
        /gene="nagD"
        /gene="nagD"
        518..1303
        /gene="nagD"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="CA06779.1"
        /db_xref="PID:e1313721"
        /db_xref="PID:g3378539"
        /db_xref="GI:3378539"
        /db_xref="SPTREMBL:O73936"
        /translation="MDYDLIISDVGVIRESEPIWENIFAIRKLKEEGKIILVTN
        NSGFSRYLTSROLNLTGLLETPNDIITSGLSAIIYKRNKRVKSVYVIGEGEIVEMSK
        NENFRVLTSEVEENNPDVAVLGLDLRLATDKLSTGMRGVARGSMFVITVMORLMPK
        DGLKLGALASSIIYALKRDPDIAGKPKKTIIOYAMELTGKIDLSKVLVIDQLET
        DIMGNDIGADTVLVLGISRKEDVENSKIKPIVVENLSLI"
        1343..3043
        /gene="sdhA"
        /gene="sdhA"
        1343..3043
        /gene="sdhA"
        /EC_number="1.3.99.1"
        /note="flavoprotein"
        /codon_start=1
        /transl_table=11
        /evidence=experimental
        /product="succinate dehydrogenase, subunit A"
        /protein_id="CA06780.1"
        /db_xref="PID:e1313722"
        /db_xref="PID:g3378540"
        /db_xref="GI:3378540"
        /db_xref="SPTREMBL:O73937"
        /translation="MEKISYDAIVIGGGLAGLAHAHEIAVSGYKAVAVISKVPTPSSH
        SSAEGIAIYIEGNSDNDNDYMTYPTIGSGDFVDPDAELLISOKSGEIVIMSW
        GLTFNRQDGRVAVRYRGQOTYRTRFVGDKLTMALHTLTFQVSGSLDDIFYENMAL
        DLIRDKRVGVAVAMEKSKMTPFFFKRAIYVATGGGKLTASTINAVINTGQIVAMA
        LRGAALKDPEYVQHFETALYPSDILISEARAGEAILINKGERFARVAPKRLDLA
        PIPTSRVITTEIRREGGFGPGVGLDRLHGLSDYIKERLALVEAKNFAVDADAFOG
        PIVRPAOHYMGIGIDVIDTIGONPDVGLGEAAEACVSHGANRLGSLNLETTIYVFG
        RETGRAVAKFLSEHKEDSNSLDKEAEKVYVDAYKFKVSSGVHFGELLINKLANTWME
        1343..3043
        /gene="sdhA"
        /gene="sdhA"
        1343..3043
        /gene="sdhA"
        /EC_number="1.3.99.1"
        /note="flavoprotein"
        /codon_start=1
        /transl_table=11
        /evidence=experimental
        /product="succinate dehydrogenase, subunit B"
        /protein_id="CA06781.1"
        /db_xref="PID:e1313723"
        /db_xref="PID:g3378541"
        /db_xref="GI:3378541"
        /db_xref="SPTREMBL:O73938"
        /translation="MSEERKEIVKIKRFSKSGSMQDEYKLVDPFTQMEVLRRIK
        TBODPLTAVRASCHMAVCGSCGMKINGEPRLACKTLALDMVKKYKNEIAIEPDPFF
        VIKDLIVDTIDYENRMFKYKPRLYPSKEVLEGGKASHRLKPEDQRELMFEOCIWGLC
        VSACPSVKNDDPEFLGPAHAKGYRFLADPROTIIDPERLKILIDSAMRCGYCYOCNVC
        PDIEPVTYIKRTAVYTKRLSEKTPVALTGEHADSIAKSIIEFTGKIDBAKYISTYG
        LITATIDMYAMQNGKIKALVYQKKVNMEDIRKIG"
        3995..4870
        /gene="sdhC"
        /gene="sdhC"
        3995..4870
        /gene="sdhC"
        /EC_number="1.3.99.1"
        /codon_start=1
        /transl_table=11
        /evidence=experimental
        /product="succinate dehydrogenase, subunit C"
        /protein_id="CA06782.1"
        /db_xref="PID:e1313724"
        /db_xref="PID:g3378542"
        /db_xref="GI:3378542"
        /db_xref="SPTREMBL:O73939"
        /translation="MKVAYPGCATGHSKVDYATKRVADYLGELVEVEDMNCGG
        GLDEYNEKAHVALNLRNLTSTYERMGDMKMTYPCVCSLOSHLALAYKYNENDLREY
        DKLKEANINYSKATAEHIVVLRDGLSEIKRAVKRPLTGLGVAGVYGGOMLRE
        OIMGPEPSNPSMEDIATGATVYKPKMATACCGFFLMGSPVAGAKIATVUNSA
        KQONADILVHPCSLCHLDQDVTQLVKVKNFNLMTLPATYITQLGLAFGSABELGI
        SKLAEVLEKKGIIIO"
        4867..5229
        /gene="sdhD"
        /gene="sdhD"
        4867..5229
        /gene="sdhD"
        /EC_number="1.3.99.1"
        /codon_start=1
        /transl_table=11
        /evidence=experimental
        /product="succinate dehydrogenase, subunit D"
        /protein_id="CA06783.1"
        /db_xref="PID:e1313725"
        /db_xref="PID:g3378543"
        /db_xref="GI:3378543"
        /db_xref="SPTREMBL:O73940"
        /translation="MSEENKIVELAKRTGATVGNVYDVSERPPRPFGEVEYKIGDI
        MNGKVHLRDGDIYHIISKIPFNKDKVKKDLKINGSIDAAGLMIKTTIPEDMYSD
        LEYIRNYLEKIKINENLKR"
        complement(5231..5598)
        /note="ORF1"
        /codon_start=3
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="CA06784.1"
        /db_xref="PID:e1313726"
        /db_xref="PID:g3378544"
        /db_xref="GI:3378544"
        /db_xref="SPTREMBL:O73941"
        /translation="DNITPYPLIKYIKNAKSKDYIKMIEYIOKINGGGGTIDISRVISA
        KYGIERDEDGIDKALSOILELRKMTSMVLDKSTTYNTEFFNALEIRMLDLATVIA
        RRLERKESRGAHYRIDYPERDDTNMLKHTIYVLAGETVEITYTKPRITIRKPRPRVY
        3045..3989
        /gene="sdhB"
        /gene="sdhB"
        3045..3989
        /gene="sdhB"
        /EC_number="1.3.99.1"
        /note="iron-sulfur protein, one [2Fe-2S] and two [4Fe-4S]
        cystein cluster motifs"
        /codon_start=1
        /transl_table=11
        /evidence=experimental
        /product="succinate dehydrogenase, subunit B"
        /protein_id="CA06781.1"
        /db_xref="PID:e1313723"
        /db_xref="PID:g3378541"
        /db_xref="GI:3378541"
        /db_xref="SPTREMBL:O73938"
        /translation="MSEERKEIVKIKRFSKSGSMQDEYKLVDPFTQMEVLRRIK
        TBODPLTAVRASCHMAVCGSCGMKINGEPRLACKTLALDMVKKYKNEIAIEPDPFF
        VIKDLIVDTIDYENRMFKYKPRLYPSKEVLEGGKASHRLKPEDQRELMFEOCIWGLC
        VSACPSVKNDDPEFLGPAHAKGYRFLADPROTIIDPERLKILIDSAMRCGYCYOCNVC
        PDIEPVTYIKRTAVYTKRLSEKTPVALTGEHADSIAKSIIEFTGKIDBAKYISTYG
        LITATIDMYAMQNGKIKALVYQKKVNMEDIRKIG"
        3995..4870
        /gene="sdhC"
        /gene="sdhC"
        3995..4870
        /gene="sdhC"
        /EC_number="1.3.99.1"
        /codon_start=1
        /transl_table=11
        /evidence=experimental
        /product="succinate dehydrogenase, subunit C"
        /protein_id="CA06782.1"
        /db_xref="PID:e1313724"
        /db_xref="PID:g3378542"
        /db_xref="GI:3378542"
        /db_xref="SPTREMBL:O73939"
        /translation="MKVAYPGCATGHSKVDYATKRVADYLGELVEVEDMNCGG
        GLDEYNEKAHVALNLRNLTSTYERMGDMKMTYPCVCSLOSHLALAYKYNENDLREY
        DKLKEANINYSKATAEHIVVLRDGLSEIKRAVKRPLTGLGVAGVYGGOMLRE
        OIMGPEPSNPSMEDIATGATVYKPKMATACCGFFLMGSPVAGAKIATVUNSA
        KQONADILVHPCSLCHLDQDVTQLVKVKNFNLMTLPATYITQLGLAFGSABELGI
        SKLAEVLEKKGIIIO"
        4867..5229
        /gene="sdhD"
        /gene="sdhD"
        4867..5229
        /gene="sdhD"
        /EC_number="1.3.99.1"
        /codon_start=1
        /transl_table=11
        /evidence=experimental
        /product="succinate dehydrogenase, subunit D"
        /protein_id="CA06783.1"
        /db_xref="PID:e1313725"
        /db_xref="PID:g3378543"
        /db_xref="GI:3378543"
        /db_xref="SPTREMBL:O73940"
        /translation="MSEENKIVELAKRTGATVGNVYDVSERPPRPFGEVEYKIGDI
        MNGKVHLRDGDIYHIISKIPFNKDKVKKDLKINGSIDAAGLMIKTTIPEDMYSD
        LEYIRNYLEKIKINENLKR"
        complement(5231..5598)
        /note="ORF1"
        /codon_start=3
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="CA06784.1"
        /db_xref="PID:e1313726"
        /db_xref="PID:g3378544"
        /db_xref="GI:3378544"
        /db_xref="SPTREMBL:O73941"
        /translation="DNITPYPLIKYIKNAKSKDYIKMIEYIOKINGGGGTIDISRVISA
```

```

BASE COUNT      1999 a      849 c      1203 g      1547 t
ORIGIN
      CEDIKGEHVKGSEVILITDGEKIAETTVRSLSKANATLISVMIRGDNALRRVSD
      TIVIVYKLLQNDLIKAYEA"

Query Match      100.0%; Score 10; DB 1; Length 5598;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGTGTGTGAC 10
      |||||
Db      3522 TGTGTGTGAC 3531

RESULT 2
AB008831      10498 bp      DNA      BCT      05-JUN-1998
LOCUS      Acinetobacter sp. genes for aniline dioxygenase subunits and
INITIATION      reductase component, partial and complete cds.
ACCESSION      AB008831
NID      92627146
VERSION      AB008831.1 GI:2627146
KEYWORDS      4-oxalocrotonate decarboxylase; 4-hydroxy-2-oxovalerate aldolase;
      acetaldehyde dehydrogenase; 2-oxopent-4-dienate hydratase;
      2-hydroxymuconic semialdehyde hydrolase; 2-hydroxymuconic
      semialdehyde dehydrogenase; catechol 2,3-dioxygenase; small
      ferredoxin-like protein; aniline dioxygenase reductase component;
      aniline dioxygenase beta-subunit; aniline dioxygenase
      alpha-subunit.
SOURCE      Acinetobacter sp. (strain:YAA) DNA.
ORGANISM      Acinetobacter sp.
      Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
      Moraxellaceae; Acinetobacter.
REFERENCE      1 (bases 1 to 10498)
AUTHORS      Takeo,M.
JOURNAL      Direct Submission
      Submitted (13-NOV-1997) to the DDBJ/EMBL/GenBank databases.
      Masahito Takeo, Himeji Institute of Technology, Department of
      Applied Chemistry; 2167 Shosha, Himeji, Hyogo 671-22, Japan
      (E-mail:takeo@chem.eng.himeji-tech.ac.jp, Tel:0792-67-4893,
      Fax:0792-67-4891)
REFERENCE      2 (sites)
AUTHORS      Takeo,M., Fujii,T., Takenaka,K., and Maeda,Y.
JOURNAL      Cloning and sequencing of a gene cluster for the meta-cleavage
      pathway of aniline degradation in Acinetobacter sp. strain YAA
      J. Ferment. Bioeng. 85, 514-517 (1998)
REFERENCE      3 (sites)
AUTHORS      Fujii,T., Takeo,M. and Maeda,Y.
JOURNAL      Plasmid-encoded genes specifying aniline oxidation from
      Acinetobacter sp. strain YAA
      Microbiology 143 (Pt 1), 93-99 (1997)
FEATURES
      location/Qualifiers
      1..10498
      /organism="Acinetobacter sp."
      /strain="YAA"
      /db_xref="taxon:472"
      1..276
      /gene="atda3"
      <1..276
      /gene="atda3"
      /codon_start=1
      /transl_table=11
      /product="aniline dioxygenase alpha-subunit"
      /protein_id="BA023550.1"
      /db_xref="PID:d1024429"
      /db_xref="PID:g2627147"
      /db_xref="GI:2627147"
      /translation="ELNALMRTOEDFPIMGSEVDVANFESQCEGLTNPETIEWIDFS
      RHNGENACYQDVIOHKPTSEIHSRHFDTWLQASAVNKENSEV"
      278..874
      /gene="atda4"

gene
CDS
      278..874
      /gene="atda4"
      /codon_start=1
      /transl_table=11
      /product="aniline dioxygenase beta-subunit"
      /protein_id="BA023551.1"
      /db_xref="PID:d1024430"
      /db_xref="PID:g2627148"
      /db_xref="GI:2627148"
      /translation="MNNKDLISIKIVDESAVNLAVYOIEIKOYSDYFNLSNGEPALD
      HKINMFLKREARLDQCEDEMTLFTLEDGCVYIPGSPAPASPASEAATPEHDIRRK
      DRIYRLQGFAYSQLPVSKTRILGAPRYMAVPPSSSEGLVTRTSITVFESDGRSQVL
      SGMGYVILKQNDLKNMKNQINLNDCLSPGNNSEFL"
      885..1895
      /gene="atda5"
      885..1895
      /gene="atda5"
      /codon_start=1
      /transl_table=11
      /product="aniline dioxygenase reductase component"
      /protein_id="BA023552.1"
      /db_xref="PID:d1024431"
      /db_xref="PID:g2627149"
      /db_xref="GI:2627149"
      /translation="MNTLKRVIDKIAETKESFVLKPLDGVLAHSPGYILPIKIR
      TEKGLFEFSYLSLSSASANEDFKITVKEKRGGRSNMCDNVKQVDFIETLPPASFH
      PONDNRQFVAFSGSGITPVISIKTALNRKRIKILFANSSESSIFRKLKQICL
      QEPRLDLOFLDEKGIPTSIARQYIDDLAEVYEGAPFPMGVEENLIESKVP
      PLITKEFASVSGDDNGDITVESSAEKQVTVNFMNLGINKSVMGSEDPFIINEIKAG
      INPSSCCAGNCSCMCLVSGDVILBSNTVLDASDEDDGVIILACRSFRSKNIETISF
      DQ"
      1902..2828
      /gene="atda8"
      1902..2828
      /gene="atda8"
      /note="similar to LysR-type transcriptional activator"
      /codon_start=1
      /transl_table=11
      /protein_id="BA023553.1"
      /db_xref="PID:d1024432"
      /db_xref="PID:g2627150"
      /db_xref="GI:2627150"
      /translation="MRCVLKHSSSYDNANMDNKAFLALYRAKDYEAASELGVDSGT
      LRRRIQSLLEVLYGSLFYKNNRNWALPGIEDVAKAVDMESATREFGILPGRGVY
      IKVSIPELVNQFSDVLAFRKIYEFETDSSDAFYDLEREGEFAIRLGRGIV
      NSLKIRGTGGLGVFGSASYIDIVDFEGDALRDSDLTKGTIGSPYKRDVPFGL
      LDWQQLGFGYNIKIMCDLDESCARFCRQAGALILPKFLAHRYFMNSCVYDASDTLAA
      ELNLISRDKSDMQVVVGNMLSAKSEMEVS"
      2888..3226
      /gene="atda6"
      2888..3226
      /gene="atda6"
      /codon_start=1
      /transl_table=11
      /product="small ferredoxin-like protein"
      /protein_id="BA023554.1"
      /db_xref="PID:d1024433"
      /db_xref="PID:g2627151"
      /db_xref="GI:2627151"
      /translation="MGESYQITTEOCSQORFPCKAGQSVLKMEDQGLECAPVGRGGG
      GGLGVTVREGDYECGRMSRVHAPPEALAOGEVLACRIYPLSLDILIECRPROSAAGLA
      NEYTTTRAMR"
      3232..4155
      /gene="atda7"
      3232..4155
      /gene="atda7"
      /codon_start=1
      /transl_table=11
      /product="catechol 2,3-dioxygenase"
      /protein_id="BA023555.1"
      /db_xref="PID:d1024434"
      /db_xref="PID:g2627152"
      /db_xref="GI:2627152"

```

```

gene
CDS
/translation="MKGVMPGHVQIRVLDVAKYIKDILGILINDRDQGRVYL
KAMTEVDKFSVVLREADEPGMDKFKCLSERVVDQLGELAAFCCELEDVEDLTKD
CGRVRVFIADPGHSEIYAIKQKGTGKMGVGNINPAMRGLEMGATFDFDLKLPN
LDETLNLFEDVYGFDLAQVAPDGRKVSQFLVSTKADHAFIHEEPGRFHNSFP
LETTQDVLAADLISMTDSTIDIGTRHGLGQIYIFDPGNNNEVFAAGDYHYPD
HEPVWDAEELGKALFYHDRVINERFNVLLT"
4187..5647
/gene="atdc"
4187..5647
/gene="atdc"
/codon_start=1
/translation="2-hydroxymuconic semialdehyde dehydrogenase"
/product_id="BAA23556.1"
/db_xref="PID:d1024435"
/db_xref="PID:g2627153"
/db_xref="GI:2627153"
/translation="MKDIRKHYINGVYGSAGSKLFDVNPANPKLISKVHEAGREVD
AAYNAKAKALGSPKCKMTLEERTSLHKNVAGINARFDEFLKAECLDGKPKSMASHI
DIPKAHFKVADMIKINVPSEETPTTALLEGVKEAGIPGVYVNVHFGSDSAGAF
MTWVGPALACANTVVPSEETPTTALLEGVKEAGIPGVYVNVHFGSDSAGAF
LTERHOVDGFTGTGTEVGEYIMKAAKIGDISLELGKNAALVFPADCDKAIIEGT
MRSAFANGGVCLGTERVYVERSTFDEFGVGLKKAEBGLKTGPDDAANIGPVLSK
HREKVLAYOKAVDPAVYGVGGVPEPELAGANVQPTITGPRDSAVITDELIF
GPCVHVCPEPDEBEAIELANSLPTGLASAISEINTRAHRAVAGQIEAGIIVNWSFMR
DLRTPFGSKOSGIGREGVHSLFETEMKNICVKL"
5655..6512
/gene="atdc"
5655..6512
/gene="atdc"
/codon_start=1
/translation="2-hydroxymuconic semialdehyde hydrolase"
/product_id="BAA23557.1"
/db_xref="PID:d1024436"
/db_xref="PID:g2627154"
/db_xref="GI:2627154"
/translation="MTAPONSPEIGRITTAAGRTNVHHDGADNAKSDVAVMMIHGSG
PGVTAMANNRLVLPBLAKNRVLAADMIGGYTERPDNTYNNRERWKAHVGVDLGG
LEQVDLVGNSFGGLALALAEHPQRRLVLMGVSGFPTKGLDGMVGPSEIEMK
RLMDVPAVKNMLITDELAEWRQASVPFQSEFAPMPAPAPROGIDNLSPEEDIR
KLPEHLLHGREDEVIPLEVLQALAEILDAAHSNVFQALWQIEHAGFAFLVNDP
LHRAODPPE"
6516..7184
/gene="atdc"
6516..7184
/gene="atdc"
/codon_start=1
/translation="2-oxopent-4-dienate hydratase"
/product_id="BAA23558.1"
/db_xref="PID:d1024437"
/db_xref="PID:g2627155"
/db_xref="GI:2627155"
/translation="MEQAKIQAFADLEYALVNRNAVPLTSRADITTEDAYHISIR
MLERQOAGARIIGIKIGVTSKAVMMNLNHDQFGYLTDMFNSGSEENISDRILA
PRAEGEIAFLIKDLTGPGVTNADVLAAECVAPCEFIIVSRILKDMKAIADODITANA
SCGLEVLGDKAVSPREVDLVTGCMVYEKNAIISTGGAALGSPVNCVTNMLNTLGO
EGIR"
7290..8198
/gene="atdc"
7290..8198
/gene="atdc"
/codon_start=1
/translation="acetaldehyde dehydrogenase"
/product_id="BAA23559.1"
/db_xref="PID:d1024438"
/db_xref="PID:g2627156"
/db_xref="GI:2627156"
/translation="MNKKIKAFIFGPGNIGTDLNKMOSSEWIEPVWVGIDPESGCL

```

Query Match

100.0%; Score 10; DB 1; Length 10498;

```

Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGTGTGTGAG 10
|||||
Db 10167 TGTGTGTGAG 10176

RESULT 3
AB016431
LOCUS
DEFINITION
ACCESSION
AB016431
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 3754)
Ohta, T., Kuroda, M. and Hayashi, H.
Direct Submission
Submitted (21-JUL-1998) to the DDBJ/EMBL/Genbank databases. Toshiko
Ohta, University of Tsukuba, College of Medical Technology;
Tennodai 1-1-1, Tsukuba, Ibaraki 305, Japan
E-mail: tohta@tsukuba.ac.jp, Tel: +81-298-53-3454,
Fax: +81-298-53-3454)
2 (sites)
Kuroda, M., Hayashi, H. and Ohta, T.
Chromosome-determined zinc responsible operon czr in Staphylococcus
aureus strain 912
JOURNAL
Microbiol. Immunol. (1999) In press
LOCATION/Qualifiers
1. 3754
/organism="Staphylococcus aureus"
/strain="912"
/db_xref="taxon:1280"
/feature="Hypothetical protein"
/translation="MFTYFKSAFKNAKPOLITLITLALAFVAVIYVLLANPOLAKY
AOTIAIYSGOGKPPVDALVYIAVILAAVSLVVOIFGITNRKRAMSHKRVK
FDLPFIARFKKGVKLSVILGLVSIAMITVLSLTLKLFSPVSEMTNS"
complement(466..471)
complement(671..716)
complement(724..1701)
/gene="czrB"
complement(724..1701)
/gene="czrB"
/function="Zinc resistance"
/feature="czrC"
/codon_start=1
/translation="2-hydroxymuconic semialdehyde dehydrogenase"
/product_id="BAA36686.1"
/db_xref="PID:d1037673"
/db_xref="PID:g4126672"
/db_xref="GI:4126672"
/translation="MSHSHHDHMSHTDNKKVLFISFLIGLFIIEIIGLLAN
SIALSDGIMHSDPFSGLVALVAFIYAKNATTKTGKYREVEALAFPGVITLFI
SLIYFAIKREFVPSVEOSKEMILISIGILVNVVAFEMFGKGDGSHNIMNAGAL
HYIGDLGSGVATITAAIILNAGRTINAPISILVYIILSANGITSSINILMEGT
HSDVDVIDEYITTKDSRIQSVHDCVHTFISDMNALSCHVAVDDITLTKCEELLEN
IEHDLILNHNMTIQLTPNKKHDESLICSGTSHSHNHAHHAAHVH"
complement(1703..2023)
/gene="czrA"
complement(1703..2023)

```

CDS

gene

```

/gene="CzrA"
/function="repressor"
/codon_start=1
/transl_table=11
/protein_id="BAA36687.1"
/db_xref="PID:d1037674"
/db_xref="PID:g4126673"
/db_xref="GI:4126673"
/translation="MSEQSEINDTLERETIFKALGDYRIRIMELLSVSEASVGH
ISHQNLSEQSNVSHQLKLSVHLVKARQSQSMIYSIDDIHVATMLKQAIHNNHFK
ESGL"
complement(1709..1714)
/gene="CzrX"
complement(2034..2039)
complement(2064..2069)
complement(2087..2092)
2161..2166
2175..2813
/feature="Hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAA36688.1"
/db_xref="PID:d1037675"
/db_xref="PID:g4126674"
/db_xref="GI:4126674"
/translation="MNLIVIGANGVGSLLVQOLAKENVPPTAGVROSQDQINLKSOG
MKAILVENDSIETLETETKPPDKVIFSVSGSGNTGADKTIIVLDGAVKSMASKE
ANIKHYVSTYDSRQAFDSDGLKPYTLAKHYADYLRRSGNLTIHFGALTNAA
GSGRIEAAQYFDGKEIPREDYATVLTKEIYVSNHNFHQEQIISGEDIDKA"
3263..3268
3275..>33754
/feature="Hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAA36689.1"
/db_xref="PID:d1037676"
/db_xref="PID:g4126675"
/db_xref="GI:4126675"
/translation="MPLFLQILKTKLWGQRLSEFGYOLDNDITGECWCYSAHPNGT
SEINGPYQGLDRLWSEHRELEGDPEPSKDFPLTKIVDARESLIHHVHNDNSYAE
HENGQYKSECCWYIIDAEDAEIYIGTIAESREYANVHQGTIESILRYIKVAPGEF"
BASE COUNT 1348 a 534 c 666 g 1206 t
ORIGIN
nucy Match 100.0%; Score 10; DB 1; Length 3754;
Jest Local Similarity 100.0%; Pred. NO. 2.7e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TGTGTGTGAG 10
| | | | | | | | | |
DB 3380 TGTGTGTGAG 3389
| | | | | | | | | |
RESULT 4
AE000810 10446 bp DNA BCT 16-APR-1998
LOCUS Methanobacterium thermoautotrophicum from bases 172512 to 182957
DEFINITION (section 16 of 148) of the complete genome.
ACCESSION AE000810 AE000666
NID 92621277
VERSION AE000810.1 GI:2621277
KEYWORDS
SOURCE
ORANISM Methanobacterium thermoautotrophicum.
Methanobacterium thermoautotrophicum.
Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
Methanobacterium.
1 (bases 1 to 10446)
Smith,D.R., Doucette-Stamm,L.A., Deloughery,C., Lee,H.-M.,
Dubois,J., Aldredge,T., Bashirzadeh,R., Blakely,D., Cook,R.,
Gilbert,K., Harrison,D., Hoang,L., Keagle,P., Lumm,W., Pothier,B.,
Qiu,D., Spadofora,R., Vicare,R., Wang,Y., Wierzbowski,J.,
Gibson,R., Jivani,N., Caruso,A., Bush,D., Safer,H., Patwell,D.,
Prabhakar,S., McDougall,S., Shiner,G., Goyal,A., Pletrovski,S.,
Church,G.M., Daniels,C.J., Mao,J.-I., Rice,P., Nolling,J. and
Reeve,J.N.
Complete genome sequence of Methanobacterium thermoautotrophicum
deltaH: functional analysis and comparative genomics
J. Bacteriol. 179 (22), 7135-7155 (1997)
96037514
2 (bases 1 to 10446)
Smith,D.R.
Direct Submission
Submitted (10-ANG-1997) Genomics and Technology Development, Genome
Therapeutics Corporation, 100 Beaver Street, Waltham, MA
02154-8448, USA
Location/Qualifiers
1..10446
/organism="Methanobacterium thermoautotrophicum"
/strain="delta H"
/db_xref="taxon:2166"
/clone="MTH"
/complement(86..796)
/gene="MTH230"
/complement(86..796)
/gene="MTH230"
/feature="Function Code:14.00 - Unknown. : similar to,
gp:GI:91652373 LN:D50905, p(-)=0.999, pid=11%"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="BA84736.1"
/db_xref="PID:g2621278"
/db_xref="GI:2621278"
/translation="MDTKNKIIIGLFIIVILGSVYENYITNSQNVGEKHYLLIC
ADGEPGPGIGAVDMAFIITMNEGINIVAVYPHNNAHTAQSAPSLRAQYSRML
HDLWENDTGERGARIAQAEIYENYNGIKTDIVVTPOAVDGILOSIPVYEGQYVS
GNSIQPLREQQQSSRGVALQSLMRAIFMATDRSKYLLMNAAGLQIQGNTIYVP
EGAVQPLISTGERVGN"
1053..1781
/gene="MTH231"
1053..1781
/gene="MTH231"
/feature="Function Code:14.01 - Unknown. Conserved protein;
similar to, pIR:LN:B64471 AC:B64471, p(-)=3.1E-32, pid=37%"
/codon_start=1
/transl_table=11
/product="conserved protein"
/protein_id="BA84737.1"
/db_xref="PID:g2621279"
/db_xref="GI:2621279"
/translation="MYSSEGAMRIYAGVGENNMERAASLAEVDLVHSEEFIEE
LRGAAAYVRSGLPAAINIMAEIKKGGPLNLSVIEVANGFLAPVLDGRVVDDE
KIAVASSEFLRTGTEPRVGVISGGRGDGRSDPEVDRSHEEPLTSMKDXRYRHH
YHLLIEAVADGCVNIIAPDITGNLIFRSVLVGTARSIGAVALGDFIVDTSSQ
TAGYLRLALKFAHWLARGMNEDNE"
1858..2625
/gene="MTH232"
1858..2625
/gene="MTH232"
/feature="Function Code:14.01 - Unknown. Conserved protein;
similar to, sp:LN:ID72_METUA AC:058767, p(-)=1.6E-72,
pid=53%"
/codon_start=1
/transl_table=11
/product="conserved protein"
/protein_id="BA84738.1"
/db_xref="PID:g2621280"
/db_xref="GI:2621280"
/translation="MSPLKPLKLYKLYEYISNLRDRMPRHVAIINDGNRRYKLOGS
MNPIDEGKGIETLEKVDWCVDLGEIYAVAFSTENRFRPEKREKGLKPRFENE
AIAENKIRHNRRVRAVAGTELLEPDVRAITIAKSTEOYSDRLVNTAIGDGOE
IYATRTKIADVAAAGLIDPEDIDEDINRRLITAGLEDPLIITRTSGEHLGSFTLMQ
SSISELFCDSVPELNRKVDFLRAINSYQORRRFV"
2682..3452
gene

```

gene  
CDS  
/gene="MTH233"  
2682. 3452  
/note="Function Code:14.01 - Unknown, Conserved protein;  
similar to, p1r:LN:E64497 AC:E64497, p()=8.6E-44, pld=368"  
/codon\_start=1  
/transl\_table=11  
/product="conserved protein"  
/protein\_id="AAB84739.1"  
/db\_xref="p1r:LN:E64497 AC:E64497, p()=8.6E-44, pld=368"  
/db\_xref="GI:2621281"  
/translation="MGKIIITIVHCHLDFKDFNRNNEETIEBARKSLRAVIDSGVGLG  
GRRRLLEASLNPGEICPTMGFHPVDASAKADLIGEVVSOIESNTDLIVAGETGMD  
FHTRDEGRROEETFRFVEELAEHEMPLVHARADAEKALETVLEAYVEVIFHC  
YGSIEIARIRLDEGVYISITLVAESEMHELVRAIPLEGMLTEDSYLSLPSFGKR  
NEPAVEAVRAIRAKMDLEDVDVSDITANMERVGL"  
complement(3518. 3895)  
/gene="MTH234"  
complement(3518. 3895)  
/note="Function Code:13.07 - Other, Unclassified; similar  
to, sp:LN:DC4C\_ACICA AC:P20370, p()=1.2E-12, pld=288"  
/codon\_start=1  
/transl\_table=11  
/product="gamma-carboxymuconolactone decarboxylase"  
/protein\_id="AAB84740.1"  
/db\_xref="p1r:LN:DC4C\_ACICA AC:P20370, p()=1.2E-12, pld=288"  
/db\_xref="GI:2621282"  
/translation="MERYRRGMEILNRMRKSYTAIRDELVEVADPLARFVAPVAGD  
VSRGVLDKTRLELTLALVLRADDDQLKSHVRLGALNKGCSKDEITIEVMIMAVYAG  
FPAINAVLAAEVFTENDPAEV"  
3989. 4639  
/gene="MTH235"  
3989. 4639  
/note="Function Code:9.02 - Metabolism of Cofactors and  
Vitamins, Riboflavin metabolism (B2); similar to,  
p1r:LN:G64383 AC:G64383, p()=3.8E-50, pld=498"  
/codon\_start=1  
/transl\_table=11  
/product="riboflavin-specific deaminase"  
/protein\_id="AAB84741.1"  
/db\_xref="p1r:LN:G64383 AC:G64383, p()=3.8E-50, pld=498"  
/db\_xref="GI:2621283"  
/translation="MRPYILNNAATLDGKIATATGSSRISGEEDLRRVHELRRECD  
IWDGINTVADDPRLVHRVDAAPGNPVRVYDGMARPPFRVNLNDEAPVIGVSE  
SAPPERVAELRRAEVVAVATRRVDLHLLERLHGMGTERLMBEGSTLNTSMITGLG  
VDEVYCIAPMTVIGGRDARTLVDSGIDEMADAIRLELKRSTLTGDELIVETVKG"  
4642. 5187  
/gene="MTH236"  
4642. 5187  
/note="Function Code:14.01 - Unknown, Conserved protein;  
similar to, p1r:LN:A64438 AC:A64438, p()=8.3E-46, pld=488"  
/codon\_start=1  
/transl\_table=11  
/product="conserved protein"  
/protein\_id="AAB84742.1"  
/db\_xref="p1r:LN:A64438 AC:A64438, p()=8.3E-46, pld=488"  
/db\_xref="GI:2621284"  
/translation="VFMLEDIYERIVIRIEGRCRECKVVCQMDFOFNQMSRLDL  
QETISRSVPVPAIDPMISTELVGRDDNDNRLIGYPCCRSFSENPVAVIDG  
EHLAVSELDIEGKCAIMPGFIPCSJRCOEAMERKLIIGADDERFRIELDEL  
NMRHPHFHAYEVEFKIYLE"  
5346. 9311  
/gene="MTH237"  
5346. 9311  
/note="Function Code:9.10 - Metabolism of Cofactors and  
Vitamins, Porphyrin and chlorophyll metabolism; similar  
to, p1r:LN:S64721 AC:S64721, p()=1E-97, pld=248"  
/codon\_start=1  
/transl\_table=11

/product="magnesium cheletase subunit"  
/protein\_id="AAB84743.1"  
/db\_xref="p1r:LN:S64721 AC:S64721, p()=1E-97, pld=248"  
/db\_xref="GI:2621285"  
/translation="WNIDKSOGENMKRFLIALILLISVPSVSAADNNTTSNTES  
KMLKDTENVALLGCVGVDPIMNEAVARNLDLPGVNFILKTYMTDFNINSVATQ  
FKNDKTVDFIIEFTFRPGYPTGMSVGFALVDLQIAASLKGAFIEVLPKPN  
ITGVVNTLPAVGPVAPALSRNKRKRVVILNLSGVNTLANTDKIYPMOQELV  
PLAPOVYDRAPIYMDMNTTLIRPGARVAVAILNRYSGNNDYETLQKLESRG  
LNVIFRYCSDPIGASRFEVWNTSVDAVACVQFQVWPNQITFTLQNLVPOG  
PLVEFLQSLDDYINSTROQLGLEVYMLNFEKQIEPILJGSHISQDPDLTV  
TIKEVAYEPGIDQVNRITLAWNLRRKANSDFKLAAMYEPSTHEDGAPANGLR  
SLSNLLMRAGSHGTGNGNTEIYELLINRAGSNPNMOTELARKVLDGNTIPV  
SEYKRWKPLASGVEALMGAPGIMINNSVYPIGMGMGILLIAPQVKKMG  
TLNNTLPTTHQFAFYXLOKGCADAVYICGHTELLPGRHNMATEDWNTLI  
GISRPHILKNEPDEAVINTKRRRAVITISYLPVWRLELGVQELSDLMGYS  
TAEASGDTDRNNSLESLTRDQVTRAGLEVRLGVNSTAFVLRORLHEYLELGLT  
PGYHTFGEPLDSETELEKIDALISFDPAANTARDEIRNLIMSASNNLSLALN  
GEIOPVARSPPVINTLETPGRNNYTPDPSSIPDPAVIGSRAEMLKRRNSG  
GRPEYVADIGDVISTGOSIAIEFYLRPIYESGALIGTEIIPLESGRRRIDY  
VISDPNRGAIPAMVYINDTIRIYVNLNSSEMYRKRKYLANKATYELVAAQ  
NTSEADNADRLARTIRGLPGADPHAGVDRLMSKDDWTPQOLATYLSYIAY  
GRDLNGVOSPRLLESLELRTVDITVWISYRAPGGCTCLRVSVTVNEVNLGTORN  
SYIARTAGTPLIRLOSITYDILAVILNVLVWQKIREGSGSALQVRLTS  
DAIVDASPDPVWRADOTFLPDSVRSQFDSALOMIARVYROAHTRLVLSQELV  
AISEMLGERSGSDGDSGTTPPGGTGSGATSGGRTGSPGLSPVSGAISMDSO  
SGAASASPQSGQERAGRSYISIPTAKSSGTQLXALNIGITGLPCLLGVGYTGPMR  
K"

gene  
CDS

/gene="MTH238"  
complement(9445. 10365)  
/note="Function Code:2.02 - Energy Metabolism, Methane  
metabolism; similar to, p1r:LN:B64382 AC:B64382,  
p()=7.3E-70, pld=318"  
/codon\_start=1  
/transl\_table=11  
/product="tungsten formylmethanofuran dehydrogenase,

Query Match  
Best Local Similarity 100.0%; Score 10; DB 1; Length 10446;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTCTCAG 10  
DB 840 TGTGTGTCTCAG 849

RESULT 5  
AE000845 16189 bp DNA BCT 16-APR-1998  
LOCUS Methanobacterium thermoautotrophicum from bases 571179 to 587367  
(section 51 of 148) of the complete genome.  
ACCESSION AE000845 AE000666  
NID 92621726  
VERSION AE000845.1 GI:2621726  
KEYWORDS Methanobacterium thermoautotrophicum.  
SOURCE Methanobacterium thermoautotrophicum  
ORGANISM Archaea: Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
Methanobacterium.

REFERENCE 1 (bases 1 to 16189)  
AUTHORS Smith,D.R., Doucette-Stamm,L.A., Deloughery,C., Lee,H.-M.,  
Dubois,J., Aldredge,T., Bashirzadeh,R., Blakey,D., Cook,R.,  
Gilbert,K., Harrison,D., Hoang,L., Keagle,P., Lum,W., Pothier,B.,  
Qiu,D., Spadafora,R., Vicare,R., Wang,Y., Wierzbowski,J.,  
Gidson,R., Jiwan,N., Caruso,A., Bush,D., Saefer,H., Patwell,D.,  
Prabhakar,S., McDougall,S., Shimer,G., Goyal,A., Pletrowski,S.,  
Church,G.M., Daniels,C.J., Mao,J.-i., Rice,P., Nolling,J. and  
Reeve,J.N.  
TITLE Complete genome sequence of Methanobacterium thermoautotrophicum  
delta: functional analysis and comparative genomics  
JOURNAL J. Bacteriol. 179 (22), 7135-7155 (1997)

MEDLINE 98037514  
 REFERENCE 2 (bases 1 to 16189)  
 AUTHORS Smith,D.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-AUG-1997) Genomics and Technology Development, Genome  
 Therapeutics Corporation, 100 Beaver Street, Waltham, MA  
 02154-8448, USA

FEATURES  
 source Location/Qualifiers  
 1..16189

gene  
 CDS  
 /organism="Methanobacterium thermoautotrophicum"  
 /strain="delta H"  
 /db\_xref="taxon:2166"  
 /clone="MTH"  
 /complement(103..1338)  
 /gene="MTH645"  
 /complement(103..1338)  
 /gene="MTH645"  
 /note="Function Code:10.12 - Metabolism of Macromolecules,  
 Degradation of proteins--peptides--and glycopeptides ;  
 similar to, sp:LN:1090\_METHA AC:057555, p(1)=1.3E-88,  
 pid=468"

gene  
 CDS  
 /codon\_start=1  
 /transl\_table=11  
 /product="collagenase"  
 /protein\_id="AAB85150.1"  
 /db\_xref="PID:92621727"  
 /db\_xref="GI:2621727"  
 /translation="MLYKIKDCFSVIEPMLSPAGDFSLMALRSGADAVYIGLEG  
 CNARADNFRVSEVAEVARADHGSRLVCNTALRDDVRLKNIPPELHSGADV  
 AVASDLAVALDAEGGLDVHVSQANITRTLTLRDGVSRALISRELISRETAG  
 IASAPVEVDFPHGALCTRAISRCILSYLGRSANCDCLOPCRKSRLVSEDEF  
 DLEVSASEAVRSTLSPADLCMEH1PELDAGVHALKIEGRGRADYATVIGVRE  
 ALDRYSAGEWFEERMLSELKVFENRFGTFSEPGNSGNISEYIKEDYGEVNV  
 YRRKVGAAARLRLMRPLRVGDEVIIQCKTGALLIQVKSQIEGSEVEAEGEGVLLV  
 EDVRPGDLYRRVKKRPS"  
 /complement(1363..2787)  
 /gene="MTH646"  
 /complement(1363..2787)  
 /gene="MTH646"  
 /note="Function Code:4.01 - Nucleotide Metabolism, Purine  
 metabolism ; similar to, sp:LN:PURL\_METHA AC:057657,  
 p(1)=6.2E-121, pid=528"

gene  
 CDS  
 /codon\_start=1  
 /transl\_table=11  
 /product="amidophosphoribosyltransferase"  
 /protein\_id="AAB85151.1"  
 /db\_xref="PID:92621728"  
 /db\_xref="GI:2621728"  
 /translation="MLGESEVRDKGIVIGYSQDKKTGVAQIYVALYALQHRQESA  
 GISTFNGNDILHRGGLVCDVNPKELEKGNVIGHRYSTGESRLENQPFWS  
 EFOGGKIAIAHNGDIINSMLELEEGHNFTDSEYICHLISREDEKRMFYS  
 IKVSEOLVGSYLVLLNODLVYVDPVGIKPLAFARGSTOIVASEVAFPIGAE  
 HVDVDPGETLHNKRGSYVANAPNTRARHCFEYVYRPPSYIDGRNRYVRNTI  
 GEALVREHPANADVVPVDPDSIPALIGSRASGIPYGEGLINRYGRTFLPCEP  
 RETAVLTKMNPISLEGRVSIIPALIGSRASGIPYGEGLINRYGRTFLPCEP  
 KSPCYGIAATKEELIASIRNVEYRIIGVDSGLYSIESIVECIGIKKGLCTGC  
 LDDDYPTLPSPDSEYEMRSCSR"  
 /complement(2803..3645)  
 /gene="MTH647"  
 /complement(2803..3645)  
 /gene="MTH647"  
 /note="Function Code:14.00 - Unknown, ; similar to,  
 pIR:LN:S61081 AC:S61081:S60817, p(1)=0.0098, pid=108"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="AAB85152.1"  
 /db\_xref="PID:92621729"  
 /db\_xref="GI:2621729"  
 /translation="MDNKGSPVDVRIIVEGASDVESYRALQVSGAKYHTISSIVP  
 TTSLEIALRAVEGADIVLTATVDQGRRELDARALRGVSHIRKMLPIYGHDEV  
 LDDPLLREIENAIIRAGLDTLTGIRSLMKRLECRKRLDEYVAENTALDENTR  
 LQGEVAGNRIEISLNGELSQLBEKFRLLGEYSKLETRSELEDEKELLTETSITDLW

gene  
 CDS

RETGPEPDLEIRIFVTDHKEPGLIGGSIAPSRBDAYEMLRINSALVTEKD  
 EESS"  
 /complement(3646..3828)  
 /gene="MTH648"  
 /complement(3646..3828)  
 /gene="MTH648"  
 /note="Function Code:10.04 - Metabolism of Macromolecules,  
 Ribosomal proteins; similar to, sp:LN:RL37\_METHA  
 AC:P54011, p(1)=2.9E-20, pid=688"

gene  
 CDS  
 /codon\_start=1  
 /transl\_table=11  
 /product="ribosomal protein L37"  
 /protein\_id="AAB85153.1"  
 /db\_xref="PID:92621730"  
 /db\_xref="GI:2621730"  
 /translation="MKGRPSFGKRNKHLHRCRCGKNSYHVRKKVCAACGFSRRI  
 RYSWONKRTGORLK"  
 /complement(3894..4139)  
 /gene="MTH649"  
 /complement(3894..4139)  
 /gene="MTH649"  
 /note="Function Code:14.01 - Unknown, Conserved protein;  
 similar to, gp:GI:9495021, p(1)=1.5E-09, pid=358"

gene  
 CDS  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved protein"  
 /protein\_id="AAB85154.1"  
 /db\_xref="PID:92621731"  
 /db\_xref="GI:2621731"  
 /translation="MIDVSSORVNPORPLDALGNSLNSPVIIKKGDRFRLVKSFD  
 LHMNLVNDASELEDEGEVTRRLGTYLIRGDNITYISP"  
 /complement(4441..4923)  
 /gene="MTH650"  
 /complement(4441..4923)  
 /gene="MTH650"  
 /note="Function Code:14.01 - Unknown, Conserved protein;  
 similar to, pIR:LN:G64478 AC:G64478, p(1)=3.3E-35, pid=448"

gene  
 CDS  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved protein"  
 /protein\_id="AAB85155.1"  
 /db\_xref="PID:92621732"  
 /db\_xref="GI:2621732"  
 /translation="MTVKIRKRYHIIKKRLREIRGOIGDYSSLPEKATVEIETDDN  
 NIIIVNGEPLNFIIDRVPTLRGALKMIDIGYVVDGAVFLNAGDVAVSPGLVD  
 APEISENVTVVVDERNRPLAVGISLSGPEMERDSGKAVKTIHHIGDAIMELEY  
 "

gene  
 CDS  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved protein"  
 /protein\_id="AAB85156.1"  
 /db\_xref="PID:92621733"  
 /db\_xref="GI:2621733"  
 /translation="MLVELNDAGNITISEDVHIGILAVGSHLENHGPALPIPTDAKI  
 ASYVALALRTGARELGVLYAASEPPYKHGIMHMDRIARELFPVARKARRLNL  
 EAAVYNGHGKNDKDDMLEEGLIETAMNRYVEIGPHAGGELSAGLIIIGIAD  
 IRLDCEIPELTPHIGMIGLKKARRENREIDAAKRCENDGVNPPDYVLCORILDAVE  
 SYISDKVELKTIQEP"

gene  
 CDS  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved protein"  
 /protein\_id="AAB85157.1"  
 /db\_xref="PID:92621734"  
 /db\_xref="GI:2621734"  
 /translation="MLVELNDAGNITISEDVHIGILAVGSHLENHGPALPIPTDAKI  
 ASYVALALRTGARELGVLYAASEPPYKHGIMHMDRIARELFPVARKARRLNL  
 EAAVYNGHGKNDKDDMLEEGLIETAMNRYVEIGPHAGGELSAGLIIIGIAD  
 IRLDCEIPELTPHIGMIGLKKARRENREIDAAKRCENDGVNPPDYVLCORILDAVE  
 SYISDKVELKTIQEP"

/transl\_table=11  
/product="ribonucleotide reductase, large subunit"  
/protein\_id="AAB85157.1"  
/db\_xref="PID:g2621734"  
/db\_xref="GI:2621734"  
/translation="MISLSTALVLEERYLLRGEQGVETPEMFRVAVASAD  
EAYDDPAAVEAFATSYVGNIEFPNSPTLNNACTPINOASACVLPVEDSIDTES  
LRDAIILHKSIGSGVFSRLPRGDIVASTVAGSVSPFRLEFVAADVIKQGGRR  
RGAMAVLHSHSPDIFSFIDAKSREGPLRNENLSVAPDEMDRVPDLINPRGEEV  
DSVSRVILKRIIVAAVRSGDPGILFEDIRINRYPTPOLGRIEATNPGSGDTIVMS  
GGPRVLELEKGPPTALIRGSGPCPGGFETCRDYDLRTREGHCLRLTHDRVIV  
MDGLFWRAGELERGRIVMDAAGEPALATFRGLRGAGROVDATVYGAARFA  
NGFIYHNGEOPILTHESCNIGSVNLSMSPSSINNEKARTIHVAHFIDNYIDVA  
STPLRVEMTLRTKIGLVGAFADMLIKGITVNSVALEVAGRVMSFISSEARA  
SMELARSGSPPEKGSIMDIQGECKRNATLTITAPGSLSIAGTSSGILEAFVS  
ETRNILGRSFHELPLFTMAGRLDKRSLLEISRGVLPVAPARIRLEFVAHEI  
DPVHVMAQAFQRYVDNAVSKTVNLPADSPADVERVFAHGLGCKGVYVYSGK  
VDEYLRPEVYSGRDMTCPR"

gene  
complement(8068. .9591)  
/gene="MTH653"  
/complement(8068. .9591)  
/note="Function Code:14.01 - Unknown, Conserved protein;  
similar to, gp:GI:g1272330 LN:MTU51624, P()=3.9E-46,  
pid=248",  
/codon\_start=1  
/transl\_table=11  
/product="conserved protein"  
/protein\_id="AAB85158.1"  
/db\_xref="PID:g2621735"  
/db\_xref="GI:2621735"

CDS  
pid=248, gp:GI:g1272330 LN:MTU51624, P()=3.9E-46,  
/codon\_start=1  
/transl\_table=11  
/product="conserved protein"  
/protein\_id="AAB85158.1"  
/db\_xref="PID:g2621735"  
/db\_xref="GI:2621735"

Query Match 100.0%; Score 10; DB 1; Length 16189;  
Best Local Similarity 100.0%; Pred. No. 2e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTCTCAG 10  
|||||||  
Db 2616 TGTGTCTCAG 2625

RESULT 6  
AE000880/c  
LOCUS  
DEFINITION Methanobacterium thermoautotrophicum from bases 1008430 to 1018501  
(section 86 of 148) of the complete genome.  
ACCESSION  
AE000880 AE000666  
g2622192  
ION  
AE000880.1 GI:2622192  
SOURCE  
Methanobacterium thermoautotrophicum.  
ORGANISM  
Methanobacterium thermoautotrophicum  
Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
Methanobacterium.  
REFERENCE  
1 (bases 1 to 10072)  
Smith,D.R., Doncette-Stamm,L.A., Deloughery,C., Lee,H.-M.,  
Dubois,J., Aldredge,T., Bashirzadeh,R., Blakely,D., Cook,R.,  
Gilbert,K., Harrison,D., Hoang,L., Keagle,P., Lumm,W., Potlter,B.,  
Olud,D., Spadafora,R., Vicare,R., Wang,Y., Wierzbowski,J.,  
Gibson,R., Jivani,N., Caruso,A., Bush,D., Safer,H., Patwell,D.,  
Prabhakar,S., McDougall,S., Shimer,G., Goyal,A., Pletrovski,S.,  
Church,G.M., Daniels,C.J., Mao,J.-I., Rice,P., Nollung,J. and  
Reeve,J.N.  
Complete genome sequence of Methanobacterium thermoautotrophicum  
delta: functional analysis and comparative genomics  
J. Bacteriol. 179 (22), 7135-7155 (1997)  
98037514  
2 (bases 1 to 10072)  
Smith,D.R.  
REFERENCE  
JOURNAL  
MEDLINE  
AUTHORS  
TITLE  
DIRECT SUBMISSION  
SUBMITTED (10-AUG-1997) Genomics and Technology Development, Genome  
Therapeutics Corporation, 100 Beaver Street, Waltham, MA  
02154-8448, USA

FEATURES  
source  
Location/Qualifiers  
1. 10072  
/organism="Methanobacterium thermoautotrophicum"  
/strain="delta H"  
/db\_xref="taxon:2166"  
/clone="MTH"  
complement(277. .1050)  
/gene="MTH1092"  
complement(277. .1050)  
/gene="MTH1092"  
/note="Function Code:12.11 - Cell Processes,  
Decarboxylation; similar to, sp:LN:YC38-CTAPA AC:P48278,  
P()=6.5E-23, pid=268"  
/codon\_start=1  
/transl\_table=11  
/product="putative membrane protein"  
/protein\_id="AAB85581.1"  
/db\_xref="PID:g2622193"  
/db\_xref="GI:2622193"  
/translation="MIMAEIEGIVTLMREKRPFRSRIYTSIVPLMLIIFGT  
GLGAARFGVPGGYKAFIYPGIIVFTSISGSVYIDROYGLKEILVAPISR  
ESMVGKALGISFASMIQAILALSFIYIGTMSPOCFQWAVIALIISMGIGGIV  
IAFTVSMGFINMSFYLPIFLSGALPRTIGEDPAMVQGAVINPLIYVADALRFT  
ILRSVLPLEVNILVITFVAVLVLAFLFNKEQNLN"  
complement(1047. .1985)  
/gene="MTH1093"  
complement(1047. .1985)  
/note="daunorubicin resistance); Function Code:12.02 -  
Cell Processes, Transport of carbohydrates organic acids  
alcohols and lipids; similar to, gp:GI:e283950:g1707742,  
P()=3.3E-67, pid=318"  
/codon\_start=1  
/transl\_table=11  
/product="ABC transporter (ATP-binding)"  
/protein\_id="AAB85582.1"  
/db\_xref="PID:g2622194"  
/db\_xref="GI:2622194"  
/translation="MDHIETEDIKKRYDEFLAVKGVNLRVEGSIYGLVGRGAKT  
TISMLCTILRPTGRTGVNGYDVARDAKRVESIGVOSRALDILGRELNEHA  
ATGVGPDYDRRIIEVELIALGDKADSVTVTSGGMRIELIGGLIHHPVLEFD  
EPTLGDPROTRESIRYIEKLNREEDVYLLTHHYEEDKDCDEVAITSHGELTAD  
SPENLKEGADITIVVDARGFHEILERKODVKEAYIMDDEVKLVVERGENLVPEI  
VNLAANNYVRVLEHPTLEDVFIKYGRGISEA"  
2082. .3494  
/gene="MTH1094"  
2082. .3494  
/gene="MTH1094"  
/note="Function Code:14.01 - Unknown, Conserved protein;  
similar to, gp:GI:g1653005 LN:D90910, P()=1.4E-98,  
pid=458"  
/codon\_start=1  
/transl\_table=11  
/product="conserved protein"  
/protein\_id="AAB85583.1"  
/db\_xref="PID:g2622195"  
/db\_xref="GI:2622195"  
/translation="MDFENIEKAPNGIKGLDMITEGPGRGHNTIYGGPGYKGFIA  
MERLGAAYGPGGYVSPFDEAMENILINPSSDLRLRDLDEGLIFEDSRRGPD  
DASISLEMLKLEDAIRRTGARVYLDKVDNLEPGRONIGLELLELRLWLMG  
GVATVTSGDYGGSPHRLLEYISDCVILHRTFEGVGTRLRIYKRVSGHGLNRY  
PFIITRGAISIPITISIDSVSRVSTGIPITDEMGVGTRLRIYKRVSGHGLNRY  
KSLLSKFAVESCRRGRCILFNSNEPADQIVRMESIGIKGEFAGDLLHSDPT  
SGLEHLVYMODLMDPNDSVIVPVGLAGAGGSPETRMENAKHLFIRLIDPLKG  
RGTSISFYSISPFATYTELKSLSDITWIVLESIRANGERYRSRLIRKRGAMHS  
SSVAEYRFTDRGLIKGGSW"  
3504. .3758  
/gene="MTH1095"  
3504. .3758  
/gene="MTH1095"  
/note="Function Code:14.01 - Unknown, Conserved protein;  
similar to, gp:GI:g1653851 LN:D90917, P()=0.000000003,  
pid=358"

gene  
CDS

6871..7461  
/gene="MTH100"  
/note="Function Code:14.01 - Unknown, Conserved protein;  
similar to, sp:LN:KTHY\_METUA AC:057741, P():=0.0000032,  
pid=198"  
/codon\_start=1  
/transl\_table=11  
/product="conserved protein"  
/protein\_id="AA085589.1"  
/db\_xref="PID:g2622201"  
/db\_xref="GI:2622201"  
/translation="MIDGLDAGAGDTHAELIRRYLERGEHYVFRSPEDDNPYGRKRA  
REALRGKGVNHIKAAIFVADIVRSMLTHWRNPPDLITSRYLMAVYLDGPUS  
SVLRRTSRVLPTEYMFPLDVSPSEELRLMEDEHMEENLDLTSTREKALRLAD  
GMVINTEDPIEDVQSRIDELIDRLDGDVYEDNVP"  
7439..8221  
/gene="MTH1101"

Query Match  
Best Local Similarity 100.0%; Score 10; DB 1; Length 10072;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

cy 1 TGTGTCTCAG 10  
|||||||  
Db 2658 TGTGTCTCAG 2649

RESULT 7  
AF006665/c  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AF006665 30100 bp DNA BCT 16-OCT-1997  
Bacillus subtilis 168 region at 182 min containing the cge gene  
cluster.  
AF006665  
g2529445  
AF006665.1 GI:2529445

Bacillus subtilis.  
Bacillus subtilis  
Bacteria: Firmicutes; Bacillus/Clostridium group;  
Bacillus/Staphylococcus group; Bacillus.  
1 (bases 1 to 30100)  
Ghim,S.-Y., Jeong,Y.-M., Choi,S.-K. and Park,S.-H.  
Sequence analysis of the 30 kb region (182') of the Bacillus  
subtilis chromosome containing the cge cluster  
unpublished (1997)  
2 (bases 1 to 30100)  
Park,S.-H., Choi,S.-K., Jeong,Y.-M. and Ghim,S.-Y.  
Direct Submission  
Submitted (04-JUN-1997) Applied Microbiology Research Division,  
KRIBB, P.O.Box 115, Yusong-Gu, Taejeon 305-600, Korea

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
SOURCE

1..30100  
/organism="Bacillus subtilis"  
/strain="168"  
/db\_xref="taxon:1423"  
/map="182 min"  
74..901  
/gene="yokB"  
74..901  
/gene="yokB"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/product="yokB"  
/protein\_id="AA081138.1"  
/db\_xref="PID:g2529446"



/db\_xref="GI:2529446"  
/translation="MKIDVSDLHINHMIMVNWNOIKMKRREIYNRLISNGGEVL  
VIAGDFTEWNOQTLMWLDEAKQYERKVFYFNHNDLYLLSKSRKYSLSIGRLNLI  
OKADKNVPLIKTETETKRYFAGVMMYLPRKIEGDFEFKGSNDSDNYILNLYN  
KVGVRAMKESMDWYETLENTQVDVFSHVPHNHPSPPEPNTCYMDVPINAKH  
WVGCHLOAEFFKDDGTFHMCIGPIPYDINDPYNVIVGEEDSYKTFELTFEPI"  
945..1136  
/gene="yokK"  
945..1136  
/gene="yokK"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/product="yokK"  
/protein\_id="AAB81139.1"  
/db\_xref="PID:g2529447"  
/db\_xref="GI:2529447"  
/translation="MENEVYVFCRKNHHLFAKNPMINTLKVISEMDCPNGBEGYHN  
WLSHGSEKREKENTMK"  
1176..1307  
/gene="yokD"  
1176..1307  
/gene="yokD"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/product="yokD"  
/protein\_id="AAB81140.1"  
/db\_xref="PID:g2529448"  
/db\_xref="GI:2529448"  
/translation="MDSREQIDWTCNECNFSMIGNSDSPSCDEIDIKPKNKILD"  
1625..1891  
/gene="yokE"  
1625..1891  
/gene="yokE"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/product="yokE"  
/protein\_id="AAB81141.1"  
/db\_xref="PID:g2529449"  
/db\_xref="GI:2529449"  
/translation="WTFNIENSSDIHEFLIAAKNDVDPFLEMINKGIPRLKNI  
DFNKLPNDDRFLAMEFDEFVLQNMILIKRFGVLFQOLIA"  
2112..2333  
/gene="yokF"  
2112..2333  
/gene="yokF"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/product="yokF"  
/protein\_id="AAB81143.1"  
/db\_xref="PID:g2529450"  
/db\_xref="GI:2529450"  
/translation="MIEIFKDTGATHDLVYHSKINTFEWVDEFDIVLSDSKELNKCYP  
VKCNPYRINCKCFPAVSSIDIFSEVANS"  
2539..2751  
/gene="yokG"  
2539..2751  
/gene="yokG"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/product="yokG"  
/protein\_id="AAB81143.1"  
/db\_xref="PID:g2529451"  
/db\_xref="GI:2529451"  
/translation="MCNRRNVITPIYEDMSKYSILHQVGRIEVFQKESQYPMFARD  
SEEDYNEKCLIMQDILKKNKVSFSF"  
complement(3022..3303)  
/gene="yokH"  
complement(3022..3303)

/gene="yokH"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/product="yokH"  
/protein\_id="AAB81144.1"  
/db\_xref="PID:g2529452"  
/db\_xref="GI:2529452"  
/translation="WVLSKSLKEVCYKKQMKPVDSDRLNLELMHEYSVSIIDLVECTG  
LSKQRIINDYVGFKSMNMGIMTAPDAIGCSIEELIYWNFKERROLTK"  
3339..3926  
/gene="yokI"  
3339..3926  
/gene="yokI"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/product="yokI"  
/protein\_id="AAB81145.1"  
/db\_xref="PID:g2529453"  
/db\_xref="GI:2529453"  
/translation="MGANNQGVFEANIEKSADOKLFFYRIKDVNPMFLKGAVSK  
NKYDFLFKGYLFPFELKSTKSKISFSEKIIKPOOKYRKAOTYRNIIIGPLFOR  
REPERKVFVHINDELTKYIAEKQIKHTYKNKYKASIPLAICEIETEVRSKKKY  
NTYVLMKLGELIKKDSRDKPLHTINTPKYGV"  
3929..4105  
/gene="yokJ"  
3929..4105  
/gene="yokJ"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/product="yokJ"  
/protein\_id="AAB81146.1"  
/db\_xref="PID:g2529454"  
/db\_xref="GI:2529454"  
/translation="MEPYRVEELKKTIVQKENSINITYTODEASNDLDEFYKOF  
AQHLEALERKAE"  
complement(4140..4565)  
/gene="orfE5"  
complement(4140..4565)  
/gene="orfE5"  
/note="orfE5: a potential ORF adjacent to the DNA  
replication terminator Terry of B. subtilis; it includes  
the right attachment site for phage Spbeta (4145-4160)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAB81147.1"  
/db\_xref="PID:g2529455"  
/db\_xref="GI:2529455"  
/translation="MPKOQTAELEKPFPHNKTLYVTGGTSGISQIVKRLMLTPKQVI  
VFESKDSKOYVMSQKAYAEDEKRLFLVGDVDRHVNQVMDVIFHARLAKQVPICE  
DHPEFAIQTNLIGQNVVEALSHRVQHVINISTDAVY"  
4935..6083  
/gene="yokK"  
4935..6083  
/gene="yokK"  
/function="unknown"  
/note="orf181 at the 3' region of B. subtilis cgeAB"  
/codon\_start=1  
/transl\_table=11  
/product="yokK"  
/protein\_id="AAB81148.1"  
/db\_xref="PID:g2529456"  
/db\_xref="GI:2529456"  
/translation="MKVPEKTLMLSTAGLLISFATSASAHYVNEBHKKVTAHETD  
PVASGDADDAIPAIVHEKHPKSKLITTKSGLVYDLDKRLKSTEPGLNVDL  
RIDPFLNGEKIDIAASNSRSEKNTLEVAIDGDKLKSIDDPNHPISITNISEYGF  
SLHSQGTAFALVYVGKQGEFEQYEIVDGGGKYTGKKVREFKLNSQTEGLVADDEY  
GNLYTAEDEDAIWKFAEPGGSGQVVDRAAGDHLADIEGLTIYAPANGKGYMAS  
SOGNNSYAMERQGRNRYVANFEITDGETSDGIDVIGFGIGPKYPYIGITVAQ  
DEGINQGVANQNFRIVSEQIAGHLGEMPLHROVNPRAKLDKSDSG"  
complement(6154..7107)



```
/protein_id="AA072062.1"
/db_xref="PID:g2415394"
/translation="MKRGLHHTAFAARDQENRFTYEVILGLRLVYKTVNPDPCGY
HFTYNGNDPDTMTTFPFQSGGTVGQARVTFVSPGSLSWKRLKESGJS
LEKTLTGKGLIFDDEDTPLAIMEADKSGKSWPDGTTTNEATGKMGVLLSD
POAHTOLTESFGYTKVAEDQIVRLASAVGVIVDVLHPEKRGVGGYVHVAV
RTKKRQAKMPLIAENHLPSESLDEREFTSVYREKGLIEFIANDDEPFMDERE
AEGTSKLPEWLEKHKRQITDILPEL"
4321..5836
/gene="yodH"
/feature="yodH"
4321..5811
/gene="yodH"
/feature="yodH"
/feature="similar to E.coli proline permease Pulp"
/citation=[2]
/codon_start=1
/transl_table=11
/product="permease"
/protein_id="AA072054.1"
/db_xref="PID:g2415386"
/db_xref="GI:2415386"
/translation="MGNGLTALITATITVTVGCGIAGDRSSPRVSEVSGGR
FGGLVWFLVGAADLYATYFLGLSTAFVGSVAFRIPYSVAIYAFYFLPLKMY
AKJHKLTLADIARENKSLASLVAIVGLMIPICQLSGIQTDLQVAGGYTN
VKEVYIISFLVALYTFESGIRKPTTAIKIDLVWIMLVNVSPLIFHNGTPTMI
DLVKEAPQMLTIPSEGPKGIPIWFTASISALFEMAHAAVGFARAKSADAVRKS
MFLPLVYILVILFELFIAFLVEDTNRILALHILIONSYGVAGCFAYATIALAS
LFIQSIATGASNFANNTKYLRLIHPVSGSKTLTYRSVYVYIGALLFQMLFPIA
LVTLQLGVSQGVQIFPAIAYSLFKMOTKEATYTGLAGLAVTFIYITQSAGITE
GFWGLANMIAVYVILNLEFVKNGSNPIEGLEFKODANPNKGA"
complement(5810..5836)
/gene="yodH"
complement(5840..7246)
/gene="ctpa"
complement(5846..7246)
/gene="ctpa"
/feature="similar to Synechocystis sp. PCC6803
carboxy-terminal processing protease precursor"
/citation=[2]
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="AA072063.1"
/db_xref="PID:g2415395"
/db_xref="GI:2415395"
/translation="MKRQLKLFVILITAVYASNLTFITGNSIILQKASAGDSKF
DKLKAIEQIKSDTYQKTDKLVDAIKGMISLDPISTYMDQEQAKSFEDTISAS
FEGIGAVERKDEILIVSRIGKSPARKAIPKPDQIKVNGSVKGMNNEAVALLR
GKGTGKLELNAGVGNIDLSIKRDTIPVETVSEKMDNNIGEIQTISETTAKEL
TDAIDSLERKAGXYILDRNGPGLMEQALITMSNLEIDKGNIMOVEYNGSKYWK
AERKRYTKPTVAVNDGTASAEIYMAALHESNVLIGETTPGKTQVTAKEYDGS
STVKLYAKMLTADGEMIRKKGIPOYKALPDIYADDTIKSGUTGNVAKA
QMKMLKLVKVNMSMTDQFVSVKQFQKREKLNETGILTGDTTKLMLELQKSLSD
NDTQMEKAIEFLKEM"
7399..8100
/gene="yodH"
/feature="yodH"
7399..8100
/gene="yodH"
/feature="yodH"
/feature="similar to E.coli biotin biosynthesis protein B10C"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yodH"
/protein_id="AA072055.1"
/db_xref="PID:g2415387"
/db_xref="GI:2415387"
/translation="MSRYLEMLSLFGVAGHPGLAFSKAVLOKAPSPDOPILDAGC
GTGQATLGHILYPTVVDKDPIMLEKAKRPNBGLAPYQAELEHPPSESSS
CVSESYLSRSLSSLOEISRYLKPFGMLIGLEALUKRPMAPRAEKOMADYGFCTL
HESEWHKLRSIFGQKTEAMSLPEMEPEPTTEMDLSOTIDPIYDITLQTHYVQW
LYSEYMGHCIFIAK"
8188..8503
/gene="yodI"
/feature="yodI"
```

```
CDS
8188..8439
/gene="yodI"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yodI"
/protein_id="AA072056.1"
/db_xref="PID:g2415388"
/db_xref="GI:2415388"
/translation="MERYHLCKNHQKAVRTTERGGRVHGRITRVRDRAVIAVWG

Query Match 100.0%; Score 10; DB 1; Length 29950;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTTGTCTGAG 10
Db 14707 TGTTGTCTGAG 14716

RESULT 9
ASARCDG ASARCDG 2259 bp DNA BCT 16-FEB-1995
LOCUS Anabaena sp. (PCC 7120) argD gene.
DEFINITION X78854
ACCESSION X78854
NID 9474180
VERSION X78854.1 GI:474180
KEYWORDS argD gene; N-acetylornithine aminotransferase.
SOURCE Anabaena sp.
ORGANISM Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
REFERENCE 1 (bases 1 to 2259)
AUTHORS Flores,E.
TITLE Direct Submission
JOURNAL Submitted (19-APR-1994) E. Flores, Instituto de Biología Vegetal,
Y Fitosintesis, Univ. de Sevilla-Csic, Facultad Biología, Apartado
1113, 41080 Sevilla, SPAIN
2 (bases 1 to 2259)
AUTHORS Floriano,B., Herrero,A. and Flores,E.
TITLE Analysis of expression of the argC and argD genes in the
JOURNAL Cyanobacterium Anabaena sp. strain PCC 7120
MEDLINE J. Bacteriol. 176 (20), 6397-6401 (1994)
FEATURES
location/Qualifiers
source 1..2259
organism="Anabaena sp."
strain="PCC7120"
db_xref="taxon:1167"
clone_lib="wild type genomic in cosmid pDUCa"
clone="pCSB335-2"
325..>1711
428..1711
/gene="argD"
428..1711
/gene="argD"
/codon_start=1
/transl_table=11
/product="N-acetylornithine aminotransferase"
/protein_id="CAA55410.1"
/db_xref="PID:g580729"
/db_xref="GI:580729"
/db_xref="SWISS-PROT:P54752"
/translation="MSLQTLIQATNPPESSGAASSPSTDSFSDASVNSTGYRPLAL
ERAGGRVMDTQGEKYLDFVAGIATCIQSHAPAVAVTRQIQKILHVSNNLYIPEQ
GELAQWIIQHSQADRVFCNSGAENEAIKLARYAHTVDLIEKPIITLNASRGR
TLATITATGQAYOKYFEDPLVPGFHVNNNDISAVEAATSEIDEDYRAAATLIEPIQ
GGGVPAPGVYFQKLRQICDPTGILINDEYQVGGKRGKMGIEIAGEVEDITSA
KGIIGGIFIGAMSKKFCDFQPGEHASTFGSNPACGVALLVCOTLERENTLQWQD
REQLNSGRLAIARYPHRLTEVRWGLINLELADIPLTADYVAKAINEGLLVP
AGPKVYRFPPLIVTEAEINIAIKLEKALATVIA"
1738..1816
/feature="putative"
terminator
BASE COUNT 692 a 477 c 501 g 589 t
```



RESULT 11  
BACCELW 2497 bp DNA BCT 10-DEC-1991  
LOCUS Bacillus subtilis N-acetylmuramoyl-L-alanine amidase (cw1B) gene,  
DEFINITION complete cds.  
M81324 M61747  
ACCESSION 9142805  
M81324.1 GI:142805  
KEYWORDS N-acetylmuramoyl-L-alanine amidase; cw1B gene; major autolysin.  
SOURCE Bacillus subtilis (strain 168) DNA.  
ORGANISM Bacillus subtilis  
Bacteria; Firmicutes; Bacillus/Clostridium group;  
Bacillus/Staphylococcus group; Bacillus.  
REFERENCE 1 (bases 1 to 2497)  
AUTHORS Kuroda, A. and Sekiguchi, J.  
TITLE Molecular cloning and sequencing of a major Bacillus subtilis  
autolysin gene  
JOURNAL J. Bacteriol. 173, 7304-7312 (1991)  
JLINE 92041629  
FEATURES  
source location/qualifiers  
1..2497  
/organism="Bacillus subtilis"  
/strain="168"  
/db\_xref="taxon:1423"  
730..734  
/gene="cw1B"  
730..2269  
/gene="cw1B"  
742..2232  
/gene="cw1B"  
/EC\_number="3.5.1.28"  
/codon\_start=1  
/transl\_table=1  
/product="N-acetylmuramoyl-L-alanine amidase"  
/protein\_id="AAA2371.1"  
/db\_xref="PID:9142806"  
/db\_xref="GI:142806"  
/translation="MRSYIKVLTMCFLGLILFPTALDANSVKVGGSNRYGTAVOIS  
KOMSTSTAVIVGGSSYADAIISAPAYOGNAPLTYNSDKLSYERTKLKEQOTN  
VIYVGTAVASNTANOIKSIGISIKRIAGSNRYDTAAVAKAMGATSKVILINGFLY  
ADAPAVIPYAKNXPILFTNKTINSNTSVTDKGISSTVYVGGTSGSNYKNTL  
PSPTRIGSNRELELANIVOKLNLSTIVSNGFSYPSDAGATLAKKQSLITLN  
GENLSTGARKIIGSKNMSNEMIGNTPAVSIVKANOQKNVGGFTIPIIDGHDQDGG  
AINGLEKEVNLDIAKRVNKLNASGALPYSNDTFYSLOERVKKASQAADLEFI  
SIHANANDSPNGSEYDPTYOANSKRLEQIQKLANLGTRDGRVKTAFYVI  
KYSKMPSVIVETAFITNASDSKRLKQAVYDKMAQAQIHGTVSYIR"  
terminator 2236..2269  
/gene="cw1B"  
/note="rho-independent terminator"  
BASE COUNT 876 a 442 c 508 g 671 t  
ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 2497;  
Best Local Similarity 100.0%; Pred. No. 3e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTCTCAG 10  
|||||||  
DB 2450 TGTGTCTCAG 2450

RESULT 12  
BACENZVMEI/c 2277 bp DNA BCT 24-MAY-1993  
LOCUS Bacillus subtilis 168 enzyme I of the PEP:phototransferase system  
DEFINITION gene, 3' end, and spore photoproduct lyase gene, complete cds.  
L08809  
ACCESSION 9289267  
NID L08809.1 GI:289267  
KEYWORDS PEP:phototransferase; enzyme I; lyase; spore photoproduct lyase.

SOURCE Bacillus subtilis (strain W168) DNA.  
ORGANISM Bacillus subtilis  
Bacteria; Firmicutes; Bacillus/Clostridium group;  
Bacillus/Staphylococcus group; Bacillus.  
REFERENCE 1 (bases 1 to 2277)  
AUTHORS Fajardo-Cavazos, P., Salazar, C. and Nicholson, M.L.  
TITLE Molecular cloning and characterization of the Bacillus subtilis  
spore photoproduct lyase (spl) gene, which is involved in repair of  
UV radiation-induced DNA damage during spore germination  
JOURNAL J. Bacteriol. 175, 1735-1744 (1993)  
JLINE 93194799  
FEATURES  
source location/qualifiers  
1..2277  
/organism="Bacillus subtilis"  
/strain="W168"  
/db\_xref="taxon:1423"  
<1..522  
/note="phototransferase system"  
/codon\_start=1  
/transl\_table=1  
/product="enzyme I"  
/protein\_id="AAA22414.1"  
/db\_xref="PID:9289268"  
/db\_xref="GI:289268"  
/translation="EFKEAKAILLEKEKLVKAGVAVSDIEVGMVEIPSTAVIAD  
FAKEDVPSIGTIDLIQITMADRNMRERYSLQPINPALRLITVIEPAHKEGV  
GCGEMAGDEIALPILGLGDEFSMAYSLIPARQISKLSKQEAESFEKILSMST  
TEEVAVFVKEFER"  
613..852  
/codon\_start=1  
/transl\_table=1  
/function="unknown"  
/product="9000 kDa protein"  
/protein\_id="AAA22415.1"  
/db\_xref="PID:9289269"  
/db\_xref="GI:289269"  
/translation="KSNQFNAGDTVYVIRPNHAAVNAHKEAEIVYHHPYHEGELS  
LFTTHPFAEDDVAVFASIEKSLYKELDIDYE"  
930..1958  
/codon\_start=1  
/transl\_table=1  
/product="spore photoproduct lyase"  
/protein\_id="AAA22416.1"  
/db\_xref="PID:9289270"  
/db\_xref="GI:289270"  
/translation="KQNPVVPOLVYIEPALLEYPLGOELDKFENNGIEIRRTSNO  
VRNIPKRNHLOQVRNASTLVIGVRLKFSKPSAEYAIPATGCMGHCHYCYQOT  
TMGSKRYIRYVVEEILDOADKYKRAPELFRFASCTSDIVGIDHLTKRAIE  
HROGSLGKRLRYTRFHYADHLDAKHNKTRPRRSINADYVTKNEPCTSPDKRIE  
AAVKAKAGIPPLGFIYAPYIYHGWEEGIRHLEFEKDALPQDVRHDITFELIOHRT  
KPAKRYIENYPPRTKLEDEKRRYKRWGRYGIGKYLYQDEEHALREALESYIDTFP  
NAKIEYFT"  
BASE COUNT 736 a 470 c 494 g 577 t  
ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 2277;  
Best Local Similarity 100.0%; Pred. No. 3e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTCTCAG 10  
|||||||  
DB 1389 TGTGTCTCAG 1380

RESULT 13  
BSU55043 8397 bp DNA BCT 09-MAY-1996  
LOCUS Bacillus subtilis plasmid pPOD2000 Rep, RepAB, RepA, ParA, ParB,  
DEFINITION and ParC genes, complete cds.  
U55043  
ACCESSION 91305508  
NID U55043.1 GI:1305508  
VERSION 055043.1

[illegible]

	Azevedo, V., Bertoletto, M.G., Bessieres, P., Bolotin, A., Borchert, S.S., Borris, R., Boutsier, L., Brans, A., Braun, M., Brignelli, S.C., Bron, S., Broillet, S., Bruschi, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Connerston, I.F., Cummings, N.J., Daniel, R.A., Dentzot, F., Devine, K.M., Dusterhoff, A., Ehrlich, S.D., Emerson, P.T., Entian, K.P., Erington, J., Fabre, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Funas, S., Galizzi, A., Galleron, N., Ghim, S.Y., Glaser, P., Gottreau, A., Goldlighty, E.J., Grandt, G., Guiseppl, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R., Henaud, A., Hilbert, H., Holstappel, S., Hosono, S., Hallio, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klauer-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidas, A., Ladinolis, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Mauel, C., Medigue, C., Medina, N., Mellado, R.P., Minato, M., Mostl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogawara, A., Outega, B., Park, S.H., Parro, V., Pohl, T.M., Portetlelle, D., Porwollit, S., Prescott, A.M., Prescan, E., Pujo, P., Punelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadate, I., Sato, T., Schenau, E., Seidel, S., Schroeter, R., Scotione, F., Sekiguchi, J., Sekowska, A., Serot, S.J., Serrot, P., Shin, B.S., Soldo, B., Sokokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takekura, K., Takeuchi, M., Tanakoshi, A., Tanaka, T., Terpirra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenberg, M., Vanlier, F., Vassaret, A., Viaril, A., Wambut, R., Wedler, E., Wedler, H., Weltzenegger, T., Winters, P., Wipat, A., Yamamoto, H., Yamaue, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and Danchin, A.
TITLE	The complete genome sequence of the gram-positive bacterium Bacillus subtilis
JOURNAL MEDLINE	Nature 390 (6657), 249-256 (1997)
REFERENCE	98044033
ADTHONS	2 (bases 1 to 213190)
ADTHONS TITLE	Direct Submissions
JOURNAL	Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Republique de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszerpasteur.fr, adanchinpasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48
FEATURES	Location/Qualifiers
source	1. 213190
	/organism="Bacillus subtilis"
	/strain="168"
	/db_xref="taxon:1423"
gene	33..740
	/gene="ydfS"
CDs	33..740
	/gene="ydfS"
	/function="unknown"
	/note="similar to hypothetical proteins"
	/codon_start=1
	/transl_table=11
	/protein_id="CAB12373.1"
	/db_xref="PID:e1182533"
	/db_xref="PID:g2632867"
	/db_xref="GI:2632867"
	/db_xref="SPTRMBL:P96697"
	/translation="MTELEAVIRTVASFGILLTAERILKQTSOMTIPFPAITATG ATAAGLAAYNSIKPHMAISFSIVLTIFLSTSKNRKRFPADPTVLNKGKI LESNNRKARTLDYLNQOKEKEIFENIEVLFALLETNGQLTVLRKPQEHVTKODMIA IYNQORPILINDIGEELIENNKNRRLTESWLBELRRDKIKVKETVAYAVLGND IYVDQRKHLSYEMPRE"
gene	complement(755..1186)
	/gene="ydfT"
	/gene="ydfT"
	complement(755..1186)
CDs	/complement(755..1186)
	/gene="ydfT"
	/function="unknown"
	/codon_start=1
	/transl_table=11
	/protein_id="CAB12374.1"
	/db_xref="PID:e1182534"

gene  
CDS  
  
/db\_xref="PID:g2632868"  
/db\_xref="GI:2632868"  
/db\_xref="SPTREMBL:P96698"  
translation="MDPEFKRMETITNEYKQDFWNLVLYKKAPHEHFDSEASTFYV  
DYQEENAVFIVEMPVYGEELTIRLSKQOLIKGITTVPFAEMEVLNERRYGEI  
EKIIIDPEAEFHILQIOLINGLIHISIPROVEYIAFNKGJ"  
complement(1199. .1441)  
/gene=ydga  
complement(1199. .1441)  
/gene=ydgc  
complement(1199. .1441)  
/function="unknown"  
/note="similar to hypothetical proteins from B. subtilis"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12375.1"  
/db\_xref="PID:e1182535"  
/db\_xref="PIR:g2632869"  
/db\_xref="GI:2632869"  
/db\_xref="SPTREMBL:P96699"  
translation="MPYOININIKINGVTONGNIDVGPTVNSHTANSKYGFANFSI  
GDLSPSSILNTNGSDSDVDODDIGNSAFISNQI"  
complement(1455. .1272)  
/gene=ydgc  
complement(1455. .1272)  
/gene=ydgb  
complement(1455. .1272)  
/function="unknown"  
/note="similar to hypothetical proteins from B. subtilis"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12376.1"  
/db\_xref="PID:e1182536"  
/db\_xref="PIR:g2632870"  
/db\_xref="GI:2632870"  
/db\_xref="SPTREMBL:P96700"  
translation="NPSTVINITYIKINSIGNSGITGDAINSPINNOKSQGTNSS  
FEDPSITSVMENFLINDPDVNDQTSIGNSDTSNVNAPIAPPILD"  
2026. .2613  
/gene=ydgc  
2026. .2613  
/gene=ydgc  
/function="unknown"  
/note="similar to hypothetical proteins"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12377.1"  
/db\_xref="PID:e1182537"  
/db\_xref="PIR:g2632871"  
/db\_xref="GI:2632871"  
/db\_xref="SPTREMBL:P96701"  
translation="MKSKGVSRRKLKAANEFSVRGFHDAAVSEIYKKGAFDPSEF  
YLTPSKIEPAEILTIDPHSRVRLTESILENGINTGVSVKRVLLAVEVFQPLDD  
KDIITGEFINPEAKOMKDLAMVKELEAEPGRGHSHSLDMETVAECLYGNIEHL  
TESFIITGICXPASLAAYVNILLITGMIPKGNDR"  
2610. .3091  
/gene=ydgd  
2610. .2954  
/gene=ydgd  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12378.1"  
/db\_xref="PID:e1182538"  
/db\_xref="PIR:g2632872"  
/db\_xref="PID:g2632872"  
/db\_xref="GI:2632872"  
/db\_xref="SPTREMBL:P96702"  
translation="MSIMKRVSILAENVLAGCIIVKSHVPFOVEHMRHYOYLFELTV  
TGILIDAGLANPAGIWMNRYAIGAGVLEVLMAGAIAHAMFRAROSIVAIQMCLCT  
IYSIMINGSYC"

terminator  
gene  
CDS  
3074. .3091  
/gene=ydgd  
3117. .3590  
/gene=ydgc  
3117. .3590

/gene="ydgE"  
 /function="unknown"  
 /note="similar to hypothetical proteins"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAB12379.1"  
 /db\_xref="PID:e1182539"  
 /db\_xref="PID:g2632873"  
 /db\_xref="GI:2632873"  
 /db\_xref="SPTRMBL:P96703"  
 /translation="MARTIDMEEFEERFARGSGVAFONITSGWTETKEPESGE  
 QAYNMTPNGEDSSNHYFNMTINDQGRMGWLTLYADPLPQKFAFYISGLIYAFAPG  
 KGLALQTLQTLDEAFARLGHVFAHNETAVALYLYQKMGVAMTNIIPMKOLC"  
 complement(3705. .5393)  
 /gene="exp2"  
 complement(3705. .3722)  
 /gene="exp2"  
 complement(3750. .5393)  
 /gene="exp2"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ATP-binding transport protein"  
 /protein\_id="CAB12380.1"  
 /db\_xref="PID:e1182540"  
 /db\_xref="PID:g2632874"  
 /db\_xref="GI:2632874"  
 /db\_xref="SWISS-Prot:P39115"  
 /translation="MKETVLTNVSIVVKQGYFKRHYNASVOGDDIGITGKNAGKS  
 LTLLHINDLAPAGQQLRKDDIKALVEOSTAAVSFADQTPAEKRLLEKHNHVLPRDH  
 QLSGGEKIKARLAKGSEDADLLIDEPTHLDEKSLQFIIQQLKHYNGVILVSHDR  
 YLFDEATKIMSLEDQTLIEFKGNSYKMFREKRLTQOREYKQKMERILEQNN  
 GLASMEKKAHOSSTKRKGFEYHVRKAPDAQIKSKRLEKELEKAKAPYPEYT  
 VRSIDTHTKGTGRLELVONTKAFGRPTLTKNANTIOHGEKVAIIIGPNSGKTLT  
 NIIGQETAGSVVPSANIGITQVPEPLPLEQPEELFENETFAKHQVQLNRH  
 LGFPAQWTEPIKHMSSGKGVKIKLMAYILIEEKDVLIDEPTHLIDPSSQLEETLS  
 VQSLTSLVMPNDKRYFLKTNKSLVINSNOIEKQNDVPSERNRETRLRKLTETEROE  
 VLGKLSLFTNDKQYKELDQAFNELFKRIKELDHODKKD"  
 complement(5428. .5450)  
 /gene="ydgF"  
 complement(5428. .7089)  
 /gene="ydgF"  
 complement(5428. .7089)  
 /gene="ydgF"  
 /gene="ydgF"

Query Match Best Local Similarity 100.0%; Score 10; DB 1; Length 213190;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 TGTGTGTCAG 10  
 |||||  
 DB 210334 TGTGTGTCAG 210343

RESULT	15				
ESUB0005	208430 bp	DNA	BCT	26-NOV-1997	
DEFINITION	Bacillus subtilis complete genome (section 5 of 21): from 802821 to 1011250.				
ACCESSION	Z99108	AL009126			
NID	92833055				
VERSION	Z99108.1	GI:2633055			
KEYWORDS	Bacillus subtilis.				
SOURCE	Bacillus subtilis				
ORGANISM	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae; Bacillus.				
REFERENCE	1 (bases 1 to 208430)				
AUTHORS	Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, J., Bertero, M.G., Bessières, P., Bolotin, A., Borchert, S., Boriss, R., Boursier, L., Brans, A., Braun, M., Bridnell, S.C., Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Connerthon, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devigne, K.M., Dusterhoft, A., Ehrlich, S.D., Emerson, P.T., Entlian, K.D., Erington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzia, A.,				



```

Galleron, N., Ghm, S.Y., Glaser, P., Goffeau, A., Golightly, E.J.,
Grandi, G., Guisepi, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R.,
Henauf, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
Klaerner-Blanchard, M., Klein, C., Kobayashi, T., Koelter, P.,
Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R.P.,
Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Paro, V., Pohl, T.M.,
Portetelle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P.,
Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
Rivollet, C., Roche, E., Roche, B., Rose, M., Sadale, Y., Sato, T.,
Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
Sekowska, A., Seror, S.J., Seror, P., Shin, B.S., Soldo, B.,
Sorokin, A., Taccori, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenhof, M., Vannier, F., Vassarotti, A.,
Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weltenegeger, T.,
Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and
Danchin, A.
The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)
98044033
2 (bases 1 to 208430)
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Direct Submission
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48
FEATURES
source
location/Qualifiers
1..208430
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
complement(5..1234)
/gene="yfnC"
complement(5..1234)
/gene="yfnC"
/function="unknown"
/note="similar to fosmidmycin resistance protein"
/codon_start=1
/transl_table=11
/protein_id="CAB12561.1"
/db_xref="PID:el182722"
/db_xref="PID:g2633056"
/db_xref="GI:2633056"
/db_xref="SPTREMBL:O06481"
/translation="MAIAALPEKTVQKRGTVVPIIIIGICHLNDSLOAVIPAMF
PILEKMSFTLQGLIATFLNMSSVMQPVWYDKRPYALPVGLTASMLGILG
LAFASFTILICVFPIGSAIFPEGSRVAVYMAAGKRGIAOSIYGVGSGOAMA
PITAIILVLCGFGAVFTLVAAALVAMFLIAKVFASRLGSLAOKSGKOKKTEMT
ATKSVSALIIIFLIFARSWYTAICNGFTFYFAMDYHVSIOQASQITVFLLFGA
IGTFLGGLADREFKRFVILGSLCSAPAIYVPRAGPLAIGVALIGLVMSSEFSV
TVYVQELVPGKIGTSGLTGLAFGMGAIAGVALIDAAGLPTMAIAFLPVLG
IIAFLPDSQKREHWS"
complement(1345..2052)
/gene="yfnB"
complement(1345..2052)
/gene="yfnB"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12562.1"
/db_xref="PID:el182723"
/db_xref="PID:g2633057"
/db_xref="GI:2633057"
/db_xref="SPTREMBL:O06480"

```

```

/translation="MKRYRTLLFDVDDPTILDFQAAEALATRLLEDONITPLNDKAO
YTTINQGLMRAEESEKMRDEYVNRFRSALLKEKYEDADGALBQTRFEEHQOLI
DFAFDLISLQOQFDLYITNGVSHQRLKRLDSGLPFKDIPESEDFQPKPKRY
FNVERLIPQFAEHTLLIGDSLTDIKGGLAGDTCWMNDMPNPPEIIPYEIR
KLEELYHILININTVSC"
complement(2110..3529)
/gene="yfnA"
complement(2110..2133)
/gene="yfnA"
complement(2144..3529)
/gene="yfnA"
/function="unknown"
/note="similar to metabolite transporter"
/codon_start=1
/transl_table=11
/protein_id="CAB12563.1"
/db_xref="PID:el182724"
/db_xref="PID:g2633058"
/db_xref="GI:2633058"
/db_xref="SPTREMBL:O06479"
/translation="WSSLFRKKPLETSLAQSRSKSLARTLSAFDULLIGCVIGTGI
FVITGVATAGGALLISFTLAGACLAFCYAEFSSSIPISGSVSYSTVTEGL
LAFILGMDLMEYVIALSAVANGSYFOSLAGFNHHPALTAGAPSMGAVFNLP
AAVILITAIYBSRGVKESTRENNYIVLMTAIIILITIVIGGVYKPNNSPMFGM
KGVLSAATVFPAYIGPDVAVSASBEVNPKNMPVGIISALAVTVYIYAVSLVIG
MPPYKLVNDPVSFALKFVGODAVAGIISVGAIGITVVALIYAOVRLTFVSRIG
GILPELFAKHVSEFPTPRNTWLTVGAAGLAFINTGLTALHVNMGGLAFTVISA
VILRKKEPEIKASFRVFPVPLISAGICLMEWYSLPGVTLSEVIAWAGLVYVE
LYLRKHSILNK"
3779..5236
/gene="yfnm"
3779..5236
/gene="yfnm"
/function="unknown"
/note="similar to benzaldehyde dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="CAB12564.1"
/db_xref="PID:el182725"
/db_xref="PID:g2633059"
/db_xref="PID:g2633059"
/db_xref="GI:2633059"
/db_xref="SPTREMBL:O06478"
/translation="MFOYEELNKQFPIGKNQEGSSPNVLENKNPTQKTFITFKRATA
DDVDEYAAALAKKMDVAVNPERKRTLEKAVYIEENEAIIYLLNEELGFRLLKA
AFETGLVANNIIEKAEATPEPRMEKILPEPIDGKNRYRYPAGVGVISPNFPEFLS
MKSVAFLPAGNGVVLKPKHEPTPIGGLILAKIENAGIIPAGLNVVYTDIAEIGDSF
VEHPVRIISFTGSKVGSYIGOLAMKFKKLELNGNSAFIYLEDADIEYANAAV
FSRTHOGQTCMSANRYLVHSIYDKFLELYQAKVESIKVADPMDPTIIGPLVNSRO
TDGLKTYEQALIEGAVPVKRGNGTIVETIILKDKYPPSIKKEELFGVSPMKR
DSDEADVIANETPFGLSGAVHTNLERGAFAKRIEGLMHWNDTITINDPNAFVG
EKQSGRLNENWSELEFTTLKWSVOHEKRSFY"
5250..6136
/gene="yfnm"
5250..6110
/gene="yfnm"
/function="unknown"
/note="similar to methyl-accepting chemotaxis protein"
/codon_start=1
/transl_table=11
/protein_id="CAB12565.1"
/db_xref="PID:el182726"
/db_xref="PID:g2633060"
/db_xref="GI:2633060"
/db_xref="SPTREMBL:O06477"
/translation="MELINTREKETADIIIDAFIKVARYINSIVODDITIGYDTEKIL
VNIAPKFFSLNVKAGDPLDGDITTDIARISNOKTSMVPRETEFPPIAARIPHDN
GRVYGCGIGTSESKKIDVASTASAVYBQTAASIDSEISNGSTGMSGSSGA
KVSSESGEADISVTVKGISDSNLLGNALIAAAGSGKGSFVVADEIRKLAAH
KMEKALEEE"
6113..6136
/gene="yfnm"
6245..6159

```

```

terminator
gene

```

CDS  
/gene="yfmr"  
6245..8134  
/gene="yfmr"  
/function="unknown"  
/note="similar to ABC transporter (ATP-binding protein)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12566.1"  
/db\_xref="PID:e1182727"  
/db\_xref="PID:g2633061"  
/db\_xref="GI:2633061"  
/db\_xref="SPTREMBL:006476"

/translation="MSILKAKENIKKTYGDKTLFDHISFHIENERIGLIGPNTGKST  
LKVVIAGLESIEEGELTKSGSVQVEFLHODPELPAGQTVLEHIIYSGESAVAKTLREYE  
KALYELGKDPENEOQROKHLAAQAKMDANNAMDANTLAKTVLSKLGVDYTKPVNELS  
GGOKKRVIAIKNLIQPADLLIDEPNHLNETIEMLEGYLSQPGAVMLVTHDRYEL  
NRYTNRYELERGSIVTYKGNVEFLEKRAEREOAEOKETRONLLRELAWLRGA  
KARSTOKARIDVETIKETGQSSSLDPAIGSHRIGROYEAENVMTAYDGRMLY  
DRNELVTPGERIGIIGPNGIGKTTLLNALAGRHIPDGDITIGQTVRIQYIYQDHS  
MNGELKVIDYIKETAENVKADDMITAEQMLERFLPRSMQOYIIRKISGGERRLY  
LLQVLMQEPNVFLDEPTNDIDETLSVLEDYIDQEPGVITVSHDRYFLDRVVDRLI  
VFEGNGVYSRFGQSYSDYMESSAKKAPPAEAEKTAEEAPKKRKLISYKQLEMD  
GIEDKIAQLEEKHEQLEADIAAGSDGKIQELMAEQAKTAEELEAAMDRTLSIMI  
EELSES"

terminator

8138..8159  
/gene="yfmr"  
/gene="yfmr"  
8257..8740  
/gene="yfmQ"  
8257..8703  
/gene="yfmQ"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12567.1"

gene  
CDS

8138..8159  
/gene="yfmr"  
/gene="yfmr"  
8257..8740  
/gene="yfmQ"  
8257..8703  
/gene="yfmQ"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12567.1"

Query Match 100.0%; Score 10; DB 1; Length 208430;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTCAG 10  
|||||  
DB 8214 TGTGTCAG 8223

Search completed: September 17, 1999, 22:01:06  
Time: 16249 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:33:30 ; Search time 425.19 Seconds

(without alignments)  
5,884 Million cell updates/sec

Title: US-08-956-518A-87

Perfect score: 10

Sequence: 1 TGTGTCTCAG 10

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

abase: N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	100.0	7377	1 Q03005	Full length T4 CDN
2	10	100.0	6151	1 Q03006	Full length T4 CDN
3	10	100.0	595	1 N90696	DNA sequence of po
4	10	100.0	591	1 N91782	DNA probe compleme
5	10	100.0	1900	1 N92624	Nucleotide sequence
6	10	100.0	5518	1 N90647	Nucleotide sequence
7	10	100.0	5566	1 N90648	Nucleotide sequence
8	10	100.0	5413	1 N90649	Nucleotide sequence
9	10	100.0	5365	1 N90646	Nucleotide sequence
10	10	100.0	1834	1 Q04690	Encodes Mammalian
11	10	100.0	7316	1 Q04555	SV40 initial promo
12	10	100.0	828	1 N81632	Human Bone Morphog
13	10	100.0	1505	1 N80632	Plasmid p170-2 enc
14	10	100.0	7533	1 N81970	Human Bone Morphog
15	10	100.0	600	1 N80979	Plasmid pEMPI-CPA.
16	10	100.0	533	1 N80239	Sequence of SV40 e
17	10	100.0	664	1 Q05513	Region between loc
18	10	100.0	646	1 Q05516	Upstream sequence
19	10	100.0	645	1 Q05517	Upstream sequence
20	10	100.0	641	1 Q05515	Upstream sequence
21	10	100.0	664	1 Q05514	Upstream sequence
22	10	100.0	6414	1 Q06283	Upstream sequence
23	10	100.0	7377	1 Q05607	Plasmid p170-2 inc
24	10	100.0	6151	1 Q05608	Plasmid p170-2 inc
25	10	100.0	560	1 Q06235	Plasmid pBG381 inc
26	10	100.0	455	1 N70071	Bovine male-specific
27	10	100.0	2400	1 N70062	5' region of human
28	10	100.0	5522	1 N70062	5' region of human
29	10	100.0	828	1 N70358	Sequence of pSV-G
30	10	100.0	7533	1 N70491	Entire sequence of
31	10	100.0	2659	1 N70216	Extracellular alka
32	10	100.0	6151	1 N90645	Sequence of T4 lym
33	10	100.0	6149	1 N90645	Sequence of T4 lym
34	10	100.0	560	1 N60553	Fragment of plasmid
35	10	100.0	1607	1 Q14036	Human BMP-2A in la
36	10	100.0	1829	1 Q14627	Human Glioblastoma
37	10	100.0	6051	1 Q14934	pBG381 sequence. N
38	10	100.0	2721	1 N50114	DNA sequence encod
39	10	100.0	782	1 N50388	Bovine interferon
40	10	100.0	782	1 N50389	Bovine interferon
41	10	100.0	840	1 N50390	Bovine interferon
42	10	100.0	840	1 N50381	Bovine interferon
43	10	100.0	6165	1 Q20324	sol. thesus CD4 seq

## ALIGNMENTS

44 10 100.0 595 1 N81528 Sequence of a port  
c 45 10 100.0 638 1 X40032 Prostate cancer as

## RESULT 1

ID 003005 standard; CDNA; 7377 BP.  
AC 003005;  
DE 29-MAY-1990 (first entry)  
DE Full length T4 CDNA of plasmid p170-2.  
KW Soluble T4; p170-2; anti-retroviral agent; AIDS; ARC; HIV; AZT.  
PN W08911860-A.  
PD 14-DEC-1989.  
PE 08-JUN-1989; U02453.  
PR 10-JUN-1988; US-204645.  
PR 20-APR-1989; US-341080.  
PA (GEHO-) General Hospital Corp.  
PI Fisher RA, Schooley RT, Hirsch MS, Johnson VA, Walker BD;  
DR WPI: 90-007302/01.  
DR P-PSDB; R04031.  
PT Combinations of soluble T4 protein and anti-retroviral agent -  
PT having synergistic activity in treatment and prevention of AIDS,  
PS Disclosure: h.  
CC Soluble T4 constructs may be produced by truncating this sequence, to  
CC remove the transmembrane and intracytoplasmic domains whilst retaining  
CC the extracellular region responsible for HIV binding. The sol. T4 is  
CC combined with an anti-viral agent such as AZT.  
SQ See also Q03006.  
SQ Sequence 7377 BP; 1760 A; 1954 C; 1902 G; 1761 T;

## Query Match

Best local Similarity 100.0%; Score 10; DB 1; Length 7377;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTCTCAG 10  
DB 22 TGTGTCTCAG 31

## RESULT 2

ID 003006 standard; CDNA; 6151 BP.  
AC 003006;  
DE 29-MAY-1990 (first entry)  
DE Full length T4 CDNA of plasmid pBG381.  
KW Soluble T4; pBG381; anti-retroviral agent; AIDS; ARC; HIV; AZT.  
PN W08911860-A.  
PD 14-DEC-1989.  
PE 08-JUN-1989; U02453.  
PR 10-JUN-1988; US-204645.  
PR 20-APR-1989; US-341080.  
PA (GEHO-) General Hospital Corp. (BIOJ) Biogen Inc.  
PI Fisher RA, Schooley RT, Hirsch MS, Johnson VA, Walker BD;  
DR WPI: 90-007302/01.  
DR P-PSDB; R04032.  
PT Combinations of soluble T4 protein and anti-retroviral agent -  
PT having synergistic activity in treatment and prevention of AIDS,  
PS Disclosure: h; 2; 100pp; English.  
CC Soluble T4 constructs may be produced by truncating this sequence, to  
CC remove the transmembrane and intracytoplasmic domains whilst retaining  
CC the extracellular region responsible for HIV binding. The sol. T4 is  
CC combined with an anti-viral agent such as AZT.  
SQ See also Q03006.  
SQ Sequence 6151 BP; 1493 A; 1589 C; 1586 G; 1483 T;

Query Match 100.0%; Score 10; DB 1; Length 6151;  
 Best Local Similarity 100.0%; Pred. No. 7.4e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTCTCAG 10  
 |||||  
 DB 26 TGTGTCTCAG 35

## RESULT 3

ID N90696 standard; DNA; 595 BP.

AC N90696;

DE 15-JUL-1990 (first entry)

DE DNA sequence of portion of plasmid pF8CIS containing SV40 enhancer and

DE promoter, cytomegalovirus splice donor site and Ig intron

KW plasmid pF8CIS; cytomegalovirus; SV40; enhancer; promoter;

KM stabilising sequence; splice donor intron sequence; Ig region;

KM splice acceptor sequence; human embryonic kidney cells (293); JM2.

Key Cytomegalovirus; SV40; +

Promoter

Location/Qualifiers

1..360

/tag= a

/note="SV40 enhancer and promoter"

361..580

/tag= b

/note="Stabilising sequence which includes

cytomegalovirus donor and intron sequence, the Ig

variable region intron and splice acceptor sequence"

480..580

/tag= c

/note="rge synthetic 100mer"

573..590

/tag= d

/note="psi-clai converter"

EP-309235-A.

22-MAR-1989.

22-SEP-1988; 308784.

25-SEP-1987; US-101712.

(GETH) Genentech Inc.

Gorman CM;

WPI: 89-095394/13.

Transient expression system for recombinant proteins -

comprising eukaryotic host cell transfected with vector

encoding trans-activating protein and expression vector

Disclosure: 44 pp; English.

Expression vector pF8CIS contains the SV40 enhancer and promoter, the

cytomegalovirus splice donor site and a portion of the intron, the Ig

intron and splice acceptor site, the cDNA encoding factor VIII and the

SV40 polyadenylation and transcription termination sites.

It is used as a method for prodn. of a

desired heterologous protein in a eukaryotic host cell.

The method comprises: transfecting a eukaryotic host cell with a vector

encoding trans-acting protein; transfecting the host cell with

an expression vector having a sequence of a ds DNA comprising, a

stabilising sequence downstream of a promoter and upstream of a DNA

encoding the amino acid sequence of the desired protein, and a poly-

adenylation sequence upstream to a transcription termination site.

The pref. host cells are human embryonic kidney cells (293) or JM2.

The method provides useful quantities of a desired protein in a

relatively short period of time without having to establish continuous

prodn. The transient expression system optimises the interaction between

specific vector components and certain trans-activating proteins.

Sequence 595 BP; 136 A; 172 C; 146 G; 141 T;

Query Match 100.0%; Score 10; DB 1; Length 595;

Best Local Similarity 100.0%; Pred. No. 6e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTCTCAG 10  
 |||||  
 DB 18 TGTGTCTCAG 27

## RESULT 4

ID N91782 standard; DNA; 591 BP.

AC N91782;

DE 16-MAR-1990 (first entry)

DE DNA probe complementary to human papilloma virus type 11

KW Cervical cancer.

OS Human papilloma virus type 11.

PN WO8909940-A.

PD 19-OCT-1989.

PE 04-APR-1989; U001318.

PR 04-APR-1988; US-177404.

PI George AL, Groff DE;

(ONCO) Oncor Inc.

DR WPI: 89-324314/44.

PT Rapid detection of specific human papilloma virus genotypes + by

PS Claim 34; page 44; 81pp; English.

CC Obd. by cutting HPV11 with BamHI and NdeI. The patent describes probes

CC (DNA or RNA) and their complements capable of detecting one or a

CC combination of HPV types 6, 11, 16, 18, 31, 33 and 35.

SO Sequence 591 BP; 148 A; 91 C; 121 G; 231 T;

Query Match 100.0%; Score 10; DB 1; Length 591;

Best Local Similarity 100.0%; Pred. No. 6e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTCTCAG 10  
 |||||  
 DB 210 TGTGTCTCAG 219

## RESULT 5

ID N92624 standard; CDNA; 1900 BP.

AC N92624;

DE 14-MAY-1990 (first entry)

DE Nucleotide sequence encoding mumps virus HN protein.

KW Haemagglutinin-neuraminidase; mumps; vaccine; ss.

OS Mumps virus.

FS Key

FT cds

Location/Qualifiers

91..1836

/tag= a

PN WO8911534-A.

PD 30-NOV-1989.

PF 17-MAY-1989; U02130.

PR 24-MAY-1988; US-198451.

PA (TEXA) Univ of Texas Syste.

PI Waxham MN;

DR WPI: 89-370727/50.

DR P-PSDB: P93666

PT HN specific degenerate oligonucleotide probe - useful for screening host

PT cells producing HN to use prod. in vaccine development

PS Claim 1; Page 25; 40pp; English.

CC Nucleotide sequence encoding HN protein of mumps virus, which may be used

CC to generate a large amounts of purified peptides from a unicellular host.

SO Sequence 1900 BP; 569 A; 433 C; 365 G; 533 T;

Query Match 100.0%; Score 10; DB 1; Length 1900;

Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTCTCAG 10  
 |||||  
 DB 1513 TGTGTCTCAG 1522

## RESULT 6

N90647

ID N90647 standard; DNA; 5518 BP.  
 AC N90647;  
 DT 01-AUG-1990 (first entry)  
 DE Nucleotide sequence of plasmid PBG396 including soluble T4-like (ST4)  
 DE polypeptide number 12 (ST4#12)  
 KM HIV; soluble T4-like polypeptide 12; immunotherapeutic; prophylactic;  
 KW plasmid PBG396; diagnostic.  
 OS Homo sapiens.  
 PN W08901940-A.  
 PD 09-MAR-1989.  
 PF 01-SEP-1988; US-02940.  
 PR 07-JAN-1988; US-141649, US-094322.  
 PA (BIOJ) Biogen Inc.  
 PI Fisher RA, Gilbert W, Sato VL, Flavell RA, Maraganore JM, Liu TR;  
 DR WPI; 89-085519/11.  
 PT DNA sequences coding for soluble T4-like polypeptide(s) -  
 PT used in immuno:therapeutic and immunosuppressive compns. and for  
 PT preventing, treating or detecting AIDS  
 PS Disclosure; ; 207pp; English.  
 CC It is the nucleotide sequence of the plasmid PBG396. The sequence  
 CC was isolated from 2 libraries: a lambda gt cDNA library derived from T  
 CC cell tumour line REX and a lambda gt10 cDNA library derived from T  
 CC peripheral blood lymphocytes (PBL). For screening a series of chemically  
 CC synthesised antisense oligonucleotide DNA probes based on the known T4  
 CC protein sequence was used. The polypeptide encoded is useful in  
 CC immunotherapeutic, prophylactic and diagnostic compns. It can be used  
 CC to purify HIV from a sample.  
 SQ Sequence 5518 BP; 1334 A; 1424 C; 1403 G; 1357 T;

Query Match 100.0%; Score 10; DB 1; Length 5518;  
 Best Local Similarity 100.0%; Pred. No. 7.4e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTGAC 10  
 |||||  
 DB 26 TGTGTGTGAC 35

RESULT 7  
 N90648  
 ID N90648 standard; DNA; 5566 BP.  
 AC N90648;  
 DT 01-AUG-1990 (first entry)  
 DE Nucleotide sequence of gene encoding mature T4 sequence of plasmid PBG393  
 DE including soluble T4-like (ST4) polypeptide number 8 (ST4#8)  
 KM HIV; soluble T4-like polypeptide 8; immunotherapeutic; prophylactic;  
 KW plasmid PBG393; diagnostic.  
 OS Homo sapiens.  
 PN W08901940-A.  
 PD 09-MAR-1989.  
 PF 01-SEP-1988; US-02940.  
 PR 07-JAN-1988; US-141649, US-094322.  
 PA (BIOJ) Biogen Inc.  
 PI Fisher RA, Gilbert W, Sato VL, Flavell RA, Maraganore JM, Liu TR;  
 DR WPI; 89-085519/11.  
 PT DNA sequences coding for soluble T4-like polypeptide(s) -  
 PT used in immuno:therapeutic and immunosuppressive compns. and for  
 PT preventing, treating or detecting AIDS  
 PS Disclosure; ; 207pp; English.  
 CC It is the nucleotide sequence of region encoding the first 182 AA of the  
 CC mature T4 sequence of plasmid PBG393. It does not contain the additional  
 CC non-T4 6 amino acids at the C-terminus following AA 182. The sequence  
 CC was isolated from 2 libraries: a lambda gt cDNA library derived from T  
 CC cell tumour line REX and a lambda gt10 cDNA library derived from T  
 CC peripheral blood lymphocytes (PBL). For screening a series of chemically  
 CC synthesised antisense oligonucleotide DNA probes based on the known T4  
 CC protein sequence was used. The polypeptide encoded is useful in  
 CC immunotherapeutic, prophylactic and diagnostic compns. It can be used  
 CC to purify HIV from a sample.  
 SQ Sequence 5566 BP; 1350 A; 1437 C; 1413 G; 1365 T; 1 Others;

Query Match 100.0%; Score 10; DB 1; Length 5566;  
 Best Local Similarity 100.0%; Pred. No. 7.4e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTGAC 10  
 |||||  
 DB 26 TGTGTGTGAC 35

RESULT 8  
 N90649  
 ID N90649 standard; DNA; 5413 BP.  
 AC N90649;  
 DT 01-AUG-1990 (first entry)  
 DE Nucleotide sequence of region encoding first 131 AA of plasmid PBG395  
 DE including soluble T4-like (ST4) polypeptide number 10 (ST4#10)  
 KM HIV; soluble T4-like polypeptide 10; immunotherapeutic; prophylactic;  
 OS Homo sapiens.  
 PN W08901940-A.  
 PD 09-MAR-1989.  
 PF 01-SEP-1988; US-02940.  
 PR 07-JAN-1988; US-141649, US-094322.  
 PA (BIOJ) Biogen Inc.  
 PI Fisher RA, Gilbert W, Sato VL, Flavell RA, Maraganore JM, Liu TR;  
 DR WPI; 89-085519/11.  
 PT DNA sequences coding for soluble T4-like polypeptide(s) -  
 PT used in immuno:therapeutic and immunosuppressive compns. and for  
 PT preventing, treating or detecting AIDS  
 PS Disclosure; ; 207pp; English.  
 CC It is the nucleotide sequence of the plasmid PBG395. The sequence  
 CC was isolated from 2 libraries: a lambda gt cDNA library derived from T  
 CC cell tumour line REX and a lambda gt10 cDNA library derived from T  
 CC peripheral blood lymphocytes (PBL). For screening a series of chemically  
 CC synthesised antisense oligonucleotide DNA probes based on the known T4  
 CC protein sequence was used. The polypeptide encoded is useful in  
 CC immunotherapeutic, prophylactic and diagnostic compns. It can be used  
 CC to purify HIV from a sample.  
 SQ Sequence 5413 BP; 1309 A; 1401 C; 1365 G; 1338 T;

Query Match 100.0%; Score 10; DB 1; Length 5413;  
 Best Local Similarity 100.0%; Pred. No. 7.4e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTGAC 10  
 |||||  
 DB 26 TGTGTGTGAC 35

RESULT 9  
 N90646  
 ID N90646 standard; DNA; 5365 BP.  
 AC N90646;  
 DT 01-FEB-1991 (first entry)  
 DE Nucleotide sequence of region encoding first 113 AA of plasmid PBG394  
 DE including soluble T4-like (ST4) polypeptide number 9 (ST4#9)  
 KM HIV; soluble T4-like polypeptide 9; immunotherapeutic; prophylactic;  
 KW plasmid PBG394; diagnostic.  
 OS Homo sapiens.  
 PN W08901940-A.  
 PD 09-MAR-1989.  
 PF 01-SEP-1988; US-02940.  
 PR 07-JAN-1988; US-141649.  
 PA (BIOJ) Biogen Inc.  
 PI Fisher RA, Gilbert W, Sato VL, Flavell RA, Maraganore JM, Liu TR;  
 DR WPI; 89-085519/11.  
 PT DNA sequences coding for soluble T4-like polypeptide(s) -  
 PT used in immuno:therapeutic and immunosuppressive compns. and for  
 PT preventing, treating or detecting AIDS  
 PS Fig 19; ; 207pp; English.  
 CC It is the nucleotide sequence of the plasmid PBG394. The sequence  
 CC was isolated from 2 libraries: a lambda gt cDNA library derived from T

CC cell tumour line REX and a lambda gt10 cDNA library derived from  
 CC peripheral blood lymphocytes (PBL). For screening, a series of chemically  
 CC synthesised antisense oligonucleotide DNA probes based on the known T4  
 CC protein sequence was used. The polypeptide encoded is useful in  
 CC immunotherapeutic, prophylactic and diagnostic compsns. It can be used  
 CC to purify HIV from a sample.  
 SQ Sequence 5365 BP; 1300 A; 1380 C; 1355 G; 1330 T;

Query Match 100.0%; Score 10; DB 1; Length 5365;  
 Best Local Similarity 100.0%; Pred. No. 7.4e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTCTCAG 10  
 |||||  
 DB 26 TGTGTGTCTCAG 35

RESULT 10  
 590

Q04690 standard; cDNA; 1834 BP.

DE 05-OCT-1990 (first entry)  
 DE Encodes Mammalian amino acid dehydrogenase activating factor-eta  
 KM Mammalian amino acid dehydrogenase activating protein-eta;  
 KM tyrosine dehydrogenase; tryptophan dehydrogenase; ss.  
 OS synthetic.

FT Key Location/Qualifiers  
 FT polyA\_signal 1671..1676

FT cds /\*tag= a  
 166..1006  
 /\*tag= b

PN J02111796-A.  
 PD 24-APR-1990.  
 PF 21-OCT-1988; 264097.  
 PR 21-OCT-1988; JP-264097.  
 PA (TOFU) Toa Nenryo Kogyo KK.  
 DR WPI: 90-169110/22.  
 DR P-PSDB; R05084.

PT Amino acid dehydrogenase activating protein -  
 PT isolated from mammal brain and activates tyrosine hydrogenase  
 PS and tryptophan hydrogenase  
 PS Disclosure: P; Japanese.

CC Probable error in specification at posn 168. C residue should read  
 CC G (to give Met codon ATG instead of ATC).  
 CC Encodes a protein characterised by Mr of 27KD (by SDS-PAGE), iso-  
 CC electric point of 4.6 and an N-terminal amino acid sequence beginning  
 MetGlyAsp.

CC Protein activates Tyr hydrogenase and Trp hydrogenase in the presence  
 CC of calcium ions and calmodulin-dependent protein phosphatase Type II.  
 CC Useful as a pharmaceutical analysis reagent.  
 CC See also Q04691 and Q04692.  
 SQ Sequence 1834 BP; 546 A; 454 C; 471 G; 363 T;

Query Match 100.0%; Score 10; DB 1; Length 1834;  
 Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTCTCAG 10  
 |||||  
 DB 1589 TGTGTGTCTCAG 1598

RESULT 11

Q04555 standard; cDNA; 7316 BP.

DT 02-OCT-1990 (first entry)  
 DE Plasmid p170.2 encoding T4 protein.  
 KM T4 protein; immunotoxin; Pseudomonas endotoxin A; AIDS; HIV; ARC;  
 KM angigenin; fusion protein; ds.  
 FH Key Location/Qualifiers

FT signal\_peptide 1199..1273  
 FT /\*tag= a  
 FT mat\_peptide 1274..2561  
 FT /\*tag= b  
 FT /product=T4 protein  
 FT replace (1457..1459,tgg)  
 FT /\*tag= c  
 FT note="differs from M12807"  
 FT replace (1958..1960,tty)  
 FT /\*tag= d  
 FT note="differs from M12807"

PN W09004414-A.  
 PD 03-MAY-1990.  
 PF 18-OCT-1989; U04584.  
 PR 18-OCT-1989; US-259355.  
 PA (BioJ) Biogen Inc.  
 PI Meade HM, Lobb RR, Gates LL, Winkler G;  
 DR WPI: 90-163876/21.  
 DR P-PSDB; R04910.

PT New immunotoxin contg. soluble T4 protein components and toxin -  
 PT esp. Pseudomonas endotoxin A, for treating or controlling AIDS  
 PT and related conditions, and new DNA sequences.

PS Disclosure: pp: English.  
 CC The T4 cDNA of p170.2 is almost identical to that reported by P.J.  
 CC Maddon et al [Cell, 42, pp 93-104 (1985)]. The Maddon sequence was  
 CC revised in 1988 to correct a DNA sequencing error at the codon for AA 3  
 CC (corrected from Asp to Lys; see M12807 in GenBank). The DNA may be  
 CC truncated (to remove transmembrane and intracellular regions) and/or  
 CC modified by SDM, pref. so the prod. extends from AAs 3-183 of the  
 CC mature protein. This DNA can then be ligated to a toxin DNA esp.  
 CC angigenin, or a fragment of pseudomonas exotoxin A contg. the  
 CC translocation and ADP-ribosylation domains. The hybrid DNA can then be  
 CC inserted into an expression vector and used to produce recombinant fusion  
 CC protein which is useful for preventing or treating AIDS, ARC, and HIV  
 CC infections. The T4 protein is an HIV receptor which binds to the virus  
 CC or to infected cells carrying gp120/160 marker antigen, so provides v.  
 CC specific targeting with minimal damage to non-target cells.  
 SQ Sequence 7316 BP; 1736 A; 1945 C; 1894 G; 1741 T;

Query Match 100.0%; Score 10; DB 1; Length 7316;  
 Best Local Similarity 100.0%; Pred. No. 7.5e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTCTCAG 10  
 |||||  
 DB 22 TGTGTGTCTCAG 31

RESULT 12

N81632 standard; DNA; 828 BP.

AC N81632;  
 DT 08-NOV-1990 (first entry)  
 DE SV40 initial promoter region.  
 KM Immunoglobulin H-chain enhancer; SV40; interferon; urokinase;  
 KM hepatitis-B surface antigen; ss.  
 PN J63160587-A.  
 PD 04-JUL-1988.  
 PF 23-DEC-1986; 305467.  
 PR 23-DEC-1986; JP-305467.  
 PA (GREC) Green Cross Corp.  
 DR WPI: 88-224895/32.

PT Expression vector used for protein prodn. - comprises combination  
 PT of immunoglobulin H-chain enhancer region and SV40 initial promoter  
 PT region.

PS Claim 3: page 2; 15pp; Japanese.  
 CC This sequence is inserted into an expression vector alongside the  
 CC human immunoglobulin heavy chain enhancer gene and the construct is  
 CC useful for the prodn. of proteins, eg interferon, urokinase or  
 CC hepatitis-B. See also N81631.  
 SQ Sequence 828 BP; 220 A; 188 C; 191 G; 229 T;

Query Match 100.0%; Score 10; DB 1; Length 828;  
 Best Local Similarity 100.0%; Pred. No. 6.2e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTGAC 10  
 |||||  
 Db 53 TGTGTGTGAC 62

RESULT 13  
 N80632 ID N80632 standard; DNA; 1606 BP.  
 AC N80632;  
 DT 08-OCT-1990 (first entry)  
 DE Human Bone Morphogenic Protein-2 class I cDNA.  
 KW Bone morphogenic protein; hBMP-2 class I; probes; cartilage formation;  
 KW bone formation; osteogenic cpds.; prodontal disease; ss.  
 Homo sapiens.  
 Key Location/Qualifiers  
 cds 356..1543  
 /tag= a  
 /product=hBMP-2 class I

MO8800205-A.  
 PD 14-JAN-1988.  
 PF 30-JUN-1987; U01537.  
 PR 26-MAR-1987; US-031346,  
 PA (GENE-) Genetics Inst Inc.  
 PI (MOZNEY JM, ROSEN VA;  
 DR WPI; 88-021565/03.  
 DR P-PSDB; P80619.  
 PT Bone morphogenic proteins - obtd. using recombinant DNA and used  
 PT for inducing cartilage and bone formation.  
 PS Disclosure: 7pp; English.  
 CC The HindIII-SacI bovine genomic hBMP-2 fragment described in N80627  
 CC is subcloned into M13, labelled and used as probe to screen  
 CC polydenylated RNAs from various cells and tissue sources.  
 CC Sequence analysis of the strongly hybridising clones hBMP-2 class I  
 CC (=BMP-2) indicated that they have extensive homology with the sequence  
 CC given in N80622.  
 CC The partial sequence is compiled from lambda U209-39 and several  
 CC other hBMP-2 class I cDNA recombinants.  
 CC This human cDNA hBMP-2 class I contains an open reading frame of 1188 bp,  
 CC encoding a protein of 396 amino acids. The protein is preceded by a  
 CC 5' untranslated region of 342 bp with stop codons in all frames. The 13  
 CC bp region preceding this 5' untranslated region represents a linker used  
 CC in the cDNA cloning procedures.  
 See also N80619-N80636 and N81963-64.  
 Sequence 1606 BP; 399 A; 430 C; 423 G; 354 T;

Query Match 100.0%; Score 10; DB 1; Length 1606;  
 Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTGAC 10  
 |||||  
 Db 14 TGTGTGTGAC 23

RESULT 14  
 N81970/C ID N81970 standard; DNA; 7533 BP.  
 AC N81970;  
 DT 12-OCT-1990 (first entry)  
 DE Plasmid pEM1-CPA.  
 KW tissue plasminogen activator; myocardial infarction; thrombosis;  
 KW embolism; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT misc\_rna 7..1748  
 FT /tag= a  
 FT /label=tpa\_cDNA

FT cds 20..1708  
 FT /tag= b  
 FT /product=t-PA  
 FT 3'utr 1749..1955  
 FT /tag= c  
 FT /label=K.3'UT  
 FT /note="from Ig kappa light chain gene"  
 FT 1930..1935  
 FT /tag= d  
 FT misc\_rna 1956..2104  
 FT /tag= e  
 FT /note="kappa 3 flanking sequence"  
 FT 2105..4463  
 FT /tag= f  
 FT /label=PBK322  
 FT /note="contg. ampicillinase gene and origin of  
 FT replication"  
 FT 4464..6971  
 FT /tag= g  
 FT /label=PSV2-gpt  
 FT /note="non-coding strand"  
 FT 6437..6547  
 FT /tag= h  
 FT /label=cap\_site  
 FT /note="from SV40(gpt)"  
 FT 6453..6469  
 FT /tag= i  
 FT /label=origin\_region  
 FT /note="from SV40"  
 FT 6548..6619  
 FT /tag= j  
 FT /rpt\_type=direct  
 FT 6620..6691  
 FT /tag= k  
 FT /rpt\_type=direct  
 FT 6633..6650  
 FT /tag= l  
 FT /note="from SV40"  
 FT 6716..6945  
 FT /tag= m  
 FT /label=blocking\_element  
 FT /note="from lambda promoter"  
 FT 6946..6971  
 FT /tag= n  
 FT /label=cap\_site  
 FT /note="lambda mRNA"  
 FT 6972..7281  
 FT /tag= o  
 FT /note="from Igh"  
 FT 7282..7466  
 FT /tag= p  
 FT /note="from metallothioneine I"  
 FT 7467..7533  
 FT /tag= q  
 FT /label=cap\_site  
 FT /note="from metallothioneine I"

MO8800242-A.  
 PD 14-JAN-1988.  
 PF 25-JUN-1987; U01569.  
 PR 26-JUN-1986; US-879038.  
 PA (DAMO-) Damon Biotech Inc.  
 PI Gillies SD;  
 DR P-PSDB; P81913.  
 PT Prod. of human tissue plasminogen activator - by myeloma cells,  
 PT pref. in presence of epsilon amino caproic acid.  
 PS Disclosure: P; English.  
 CC The plasmid was prep'd by cleaving the very long 3' UT of t-PA cDNA  
 CC (which causes mRNA instability) 34 nucleotides downstream of stop  
 CC codon and inserting it into pEM1 vector. The pEM1 vector was  
 CC constructed from the following components (see feature table):  
 CC (a) a 2.25 PvuII-BamHI fragment from PSV2-gpt contg. the SV40  
 CC enhancer and early region promoter, the E.coli gpt gene, the SV40

RESULT	15
N80979	
ID	N80979 standard; DNA: 600 BP.
AC	N80979;
DT	12-SEP-1990 (first entry)
DE	Sequence of SV40 early promoter, polylinker and SV40 poly A addition
DE	region from pSV40
KKK	Placetal-derived growth factor (PDGF): SV40 virus; pSV7d;
KKK	wound treatment; SV40 polylinker; SV40 poly A addition; ds.
OSK	SV40 virus.

	Location/Qualifiers
key	1..315
promoter	/*tag= a
	316..SV40 early promoter"
	/note=.579
misc_feature	/*tag= b
	/note="SV40 polylinker"
	580
misc_feature	/*tag= c
	/note="base 4210 of pBR322"
	580..600
misc_feature	/*tag= d
	/note="pBR322 DNA"

EP-288307-A.  
26-OCT-1988.  
22-APR-1988; 303658.  
22-APR-1987; US-041299.  
(CHIR-) Chiron Corp.  
Haffin C-H, Betsholtz C, Westermarck B, Knott TJ, Bell GI,  
Rall L;  
MPI: 88-30127/43.  
Recombinant prodn. of growth factor A-chain polypeptide -  
for use in treating wounds and ulcers caused by diseases,  
infections or neoplasms  
PS Example: Fig 7; 31pp; English.  
The mammalian cell shuttle vector plasmid pSV7d contains the SV40 origin  
of replication and early promoter (315 bp, PvuII pos 272-StuI pos 5193  
CC with an 8 bp deletion between nucleotides 173 and 182), a polylinker,  
CC and the SV40 poly A addition site (217 bp BclI pos 2775-pM 2558).  
CC The SV40 sequences were cloned into the pBR322 derivative pML between  
CC nucleotide 4210 and NruI pos 973. The SV40 sequences are positioned  
CC such that the direction of transcription from the early promoter is in  
CC the same direction as the ampicillin gene of the vector. Three separate  
CC Plasmid-derived growth factor mammalian cell expression vectors were  
CC constructed by isolating EcoRI fragments from cDNA clones and ligating  
CC them into pSV7d previously digested with EcoRI and treated with alkaline  
CC phosphatase. The resulting clones were called pSV7d-PDGF-A103 (D1),  
CC pSV7d-PDGF-A102 (A3-1), and pSV7d-PDGF-B1 (B-chain). Recombinant PDGF is  
CC applied topically to wounds such as cutaneous, dermal, mucosal or

CC	epithelial wounds.	It may be used to treat any type of full or partial
CC	thickness wounds.	
Sequence	600 BP;	166 A; 158 C; 118 G; 158 T;
Query Match		100.0%; Score 10; DB 1; Length 600;
Best Local Similarity		100.0%; Prid. NO. 6e+02;
Matches 10; Conservative		0; Mismatches 0; Indels 0; Gaps
QY	1 TGTGTGTCAG 10	
Db	9 TGTGTGTCAG 18	

Search completed: September 18, 1999, 00:33:31  
Job time: 18971 sec







```

/transl_table=11
/protein_id="BA19259.1"
/db_xref="PID:d1020012"
/db_xref="PID:g1881232"
/db_xref="GI:1881232"
/transl_table="METCKVNEHITRILPEPRDAERLAELIIONOORLKGWLEPAENP
SADTYRETIIDPMWROYADLNGIEAGLLYDGSICGMSILNIDQVNRKATIGYIAK
EPEKGITACRKLITAFEELEINRAITCAVANGENSRVPERIGLEBGRKADGL
YVNGHMDLVITSLKREGEK"
gene
6846..7284
/gene="ydaC"
RBS
6846..6855
/gene="ydaG"
6862..7284
/gene="ydaC"
/note="FUNCTION UNKNOWN."
/codon_start=1
/transl_table=11
/protein_id="BA19260.1"
/db_xref="PID:d1020013"
/db_xref="PID:g1881233"
/db_xref="GI:1881233"
/transl_table="MNOODIKQVLDVLDHKKVGSLATVOKGPHSRMPEFHDGLTI
YPTSKETHKAEIENNPNVHILIGYDEGSDGAVVEYAGAKAINNSLKDKITMSK
LERWFDGDDPNVILEFEPEDILNMGERTPVSLE"
gene
7778..8599
/gene="ydaH"
RBS
7778..7783
/gene="ydaH"
7790..8599
/gene="ydaH"
/note="FUNCTION UNKNOWN."

Query Match      100.0%; Score 11; DB 1; Length 14068;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTGTTCTACT 11
      1111111111
Db 108010 CTGTTCTACT 108000

RESULT 2
AB008030/c      6908 bp      DNA      circular      BCT      13-FEB-1999
LOCUS      Vibrio alginolyticus genes for Na-translocating NADH-quinone
DEFINITION      Reductase complex, nqr operon, complete genome.
SESSION      AB008030 D43958 D43958
NID      92558472
VERSION      AB008030.1 GI:2558472
KEYWORDS      Nqr1 subunit of Na-translocating NADH-quinone reductase complex
Alpha-subunit; Nqr2 subunit of Na-translocating NADH-quinone
reductase complex; Nqr3 subunit of Na-translocating NADH-quinone
reductase complex gamma-subunit; Nqr4 subunit of Na-translocating
NADH-quinone reductase complex; Nqr5 subunit of Na-translocating
NADH-quinone reductase complex; Nqr6 subunit of Na-translocating
NADH-quinone reductase complex beta-subunit.
SOURCE      Vibrio alginolyticus DNA.
ORGANISM      Vibrio alginolyticus
Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE      1 (bases 1 to 6908)
AUTHORS      Hayashi,M., Unemoto,T. and Sugiyama,A.
TITLE      Direct Submissions
JOURNAL      Submitted (09-OCT-1997) to the DDBJ/EMBL/Genbank databases. Mak1
Laboratory of Membrane Biochemistry, 1-33 Yayoi-cho, Inage-ku,
Chiba, Chiba 265, Japan (E-mail:makihaep.chiba-u.ac.jp,
Tel:043-290-2932, Fax:043-290-3021)
REFERENCE      2 (sites)
AUTHORS      Hayashi,M., Hirai,K. and Unemoto,T.
TITLE      Sequencing and the alignment of structural genes in the nqr operon
encoding the Na(+)-translocating NADH-quinone reductase from Vibrio
alginolyticus

```

```

JOURNAL      FEBS Lett. 363 (1-2), 75-77 (1995)
MEDLINE      95246889
REFERENCE      3 (sites)
AUTHORS      Hayashi,M., Hirai,K. and Unemoto,T.
TITLE      Cloning of the Na(+)-translocating NADH-quinone reductase gene from
the marine bacterium Vibrio alginolyticus and the expression of the
beta-subunit in Escherichia coli
JOURNAL      FEBS Lett. 356 (2-3), 330-332 (1994)
MEDLINE      95104444
REFERENCE      4 (sites)
AUTHORS      Hayashi,M., Nakayama,Y. and Unemoto,T.
TITLE      Existence of Na+-translocating NADH-quinone reductase in
Haemophilus influenzae
JOURNAL      FEBS Lett. 381 (3), 174-176 (1996)
MEDLINE      96176316
REFERENCE      5 (sites)
AUTHORS      Nakayama,Y., Hayashi,M. and Unemoto,T.
TITLE      Identification of six subunits constituting Na+-translocating
NADH-quinone reductase from the marine Vibrio alginolyticus
JOURNAL      FEBS Lett. 422 (2), 240-242 (1998)
MEDLINE      98149659
COMMENT      On Oct 22, 1997 this sequence version replaced gi:677955 gi:893412.
D43958: submitted (17-Feb-1998)
D43964: submitted (17-Feb-1995).

FEATURES
source
1..6908
/organism="Vibrio alginolyticus"
/db_xref="taxon:663"
/note="There are conflicts between this sequence and
GenBank Accession Number:Z37111."
386..1726
/gene="nqr1"
386..1726
/gene="nqr1"
/note="contains FMN as a cofactor; nqrA, Nqr1 subunit of
Na-NQR complex alpha-subunit"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="Nqr1 subunit of Na-translocating NADH-quinone
reductase complex alpha-subunit"
/protein_id="BAA22910.1"
/db_xref="PID:d1023781"
/db_xref="PID:g2558473"
/db_xref="GI:2558473"
/transl_table="MTTKKGLDLPFAGTSPQYVINGKTKIKKALLGEEYVGRPMH
VRGDEYKKAQVFEEDKKNPQVAFAPAGKVEVNRGAKVQSVIVIEAGEEOVTE
DKFEAQLSGLDREVIKTLQVDSGLTALTREPSTKPALESSTKAFVYADNPPLA
AKPELLINEQOEAFIAGLIDLSALTEGKYVCKSTLPSQSQSYEEHFVDDPHFAG
LACTHMEFLYPVNAENVAMSIYQDVIAPEKLELGEIYDRVSLAGPVPNNPLVR
LTGASLIDLTDLNEMLPGEYRISGVLTGCTHATGCHAYIGRHOQSVYLRERREKL
FGAMGKKNKSTRTSFLGIVFGQLPNMTTNGSDRSKVPIGNTERVAPLMEPTL
LRLDLGADTDSQAALGALDELDEDLALCTVCPGKYEYGTLLRECLDIYKRG"
1732..2976
/gene="nqr2"
1732..2976
/gene="nqr2"
/note="hydrophobic membrane protein with pI 8.14; nqrB,
Nqr2 subunit of Na-NQR complex"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="Nqr2 subunit of Na-translocating NADH-quinone
reductase complex"
/protein_id="BAA22911.1"
/db_xref="PID:d1023782"
/db_xref="PID:g2558474"
/db_xref="GI:2558474"
/transl_table="MALKKFLDIEHHPFGCKHEKFPALYENAVATFVTPGIVTKS
SHRSDVDLKRITIVYFAVFPFMGMNAGQAIALNHNITAGDQLTAVISGNHY
WTEMLGGITADAGVSKMLGATVFLIVATVFLVGGFWELFCVNRKHEVNEGF
VILSLFALIVPPLPLMOAALGITFGVVAKEIFGCTGGRNFLPALAGRAFLFFAYPA
QISGDVVTMTAADFGATALSQAQGNAGALVNTVYGSPIYMDAFIGNIPSGIEVS

```

TLALMGAMVYRIASRIAGVMTGMAVSTLNVIGSDINPFMFHMLVIG  
GFAFDFMFATDPVASFSTKNGKWYSILIGAMCMVIRVNPAYPEGMALIFANLE  
APFDHVIENKIRRLARYGK"  
/db\_xref="PID:d1023786"  
/db\_xref="PID:g2558478"  
/db\_xref="GI:2558478"  
/gene="ngq3"  
2966. 3736  
/gene="ngq3"  
/note="acidic protein with pI 4.83; ngq3, Ngq3 subunit of  
Na-NQR complex gamma-subunit"  
/codon\_start=1  
/transl\_table=11  
/evidence="experimental"  
/product="Ngq3 subunit of Na-translocating NADH-quinone  
reductase complex gamma-subunit"  
/protein\_id="BAA22912.1"  
/db\_xref="PID:d1023783"  
/db\_xref="PID:g2558475"  
/db\_xref="GI:2558475"  
/translation="MASNDSIKTKLGLVIGLSVCSIIYSTAAVGLRDKANAVLD  
KQSRIVEAGIDANGKVPPELFAVIEPRVLDLETGFTSGNASTYDQRAKDAENS  
IALTPEDVDADIRRRANAVVYLVKDDDEVKILLPHGGLMSMTAFYAVEIDGNT  
VSAITYYEOGEIPGLGEVENSMDQFIEGKILYNEDHQAIAKVKAGAPQSGSHGVD  
GLSGATLTSNGVHTFDFWLGSDKGFPLAKVRGELN"  
3736. 4368  
/gene="ngq4"  
3736. 4368  
/gene="ngq4"  
/note="hydrophobic membrane protein with pI 9.17; ngq4,  
Ngq4 subunit of Na-NQR complex"  
/codon\_start=1  
/transl\_table=11  
/evidence="experimental"  
/product="Ngq4 subunit of Na-translocating NADH-quinone  
reductase complex"  
/protein\_id="BAA22913.1"  
/db\_xref="PID:d1023784"  
/db\_xref="PID:g2558476"  
/db\_xref="GI:2558476"  
/translation="MSSQNVKSKSLAPVLDNNPIALQVLGVCSALAVTKLEAFVY  
TLAVFTVALSNFSLIRHNPNSVRIYOMALIASIVYDQVLKAYLSDISKOLS  
VFGLITINCIVMGRAEAFKAPSAVPSLIDGNGCGYGVLTGFFPEELSGSL  
FGLVLPVSGKGYQPRGMLLAPSAFFLIGFLIWIYRIKPEQVZAKE"  
4376. 4972  
/gene="ngq5"  
4376. 4972  
/gene="ngq5"  
/note="hydrophobic membrane protein with pI 7.00; ngq5,  
Ngq5 subunit of Na-NQR complex"  
/codon\_start=1  
/transl\_table=11  
/evidence="experimental"  
/product="Ngq5 subunit of Na-translocating NADH-quinone  
reductase complex"  
/protein\_id="BAA22914.1"  
/db\_xref="PID:d1023785"  
/db\_xref="PID:g2558477"  
/db\_xref="GI:2558477"  
/translation="MEHYISLVKSIETENALSFIFGCTFLAVSKRYTSFGLVIA  
VVVLTAVPNVNIYNDVRENALVGVDSIFNFTFIFGIALVQILLENVDREF  
PPLYNALGIFLPLITVNCALFGVSEVQORDNFAESIYGFSGVGVMALAIYALAI  
REKMKSDVPEGLRGRLGITFTVGLMALGFMSFSGVOL"  
5006. 6229  
/gene="ngq6"  
5006. 6229  
/gene="ngq6"  
/note="contains FAD as a cofactor; NADH dehydrogenase  
activity, pI 4.55; ngq6, Ngq6 subunit of Na-NQR complex  
beta-subunit"  
/codon\_start=1  
/transl\_table=11  
/evidence="experimental"  
/product="Ngq6 subunit of Na-translocating NADH-quinone  
reductase complex beta-subunit"

/protein\_id="BAA22915.1"  
/db\_xref="PID:d1023786"  
/db\_xref="PID:g2558478"  
/db\_xref="GI:2558478"  
/translation="MDIIGVWFTLIVALLVILVIFAKSKIVPTGDIITISVNDPSL  
AATPGGKILSLAAGAVYVSACGGGSCCGCRKAVSGSGDIITPTEDHTKGEA  
RSEERLACQVAMKTDMDIPEPTEFVKKWECTVISDNKATITKEKLDIPGESVP  
FRAGYIQIEAPAHVYADYDIPEDYREDWENFNLFRYSKYNEETIRAYSNANTPE  
EHGIIINLVRIATPPNPDPVPPGISYIWSKEDDKTISGPFEPFAKDTDAEMV  
FVGGAGAMPARSHIFDOLKRLSKRMSFVGARSRKREMYEDFDMQAEKDNFVW  
HCAISDPLPEDNMDGYTGFIHNLVYENYLRDHEAPDCEYMGCPMMAAVIGMLKD  
ICVEDNIIIDDDG"  
terminator 6252. 6276  
BASE COUNT 1840 a 1473 c 1619 g 1976 t  
ORIGIN  
Query Match 100.0%; Score 11; DB 1; Length 6908;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CTGTTCTAGT 11  
|||||||  
Db 3856 CTGTTCTAGT 3846  
RESULT 3  
AB008698  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
1. 1167  
/organism="Acinetobacter radiorestrictens"  
/strain="SEIP 12.81"  
/db\_xref="taxon:40216"  
1. 1167  
/gene="gyrB"  
/gene="gyrB"  
/gene="gyrB"  
/BC\_number="5.99.1.3"  
/codon\_start=1  
/transl\_table=11  
/product="DNA gyrase subunit B"  
/protein\_id="BAA75415.1"  
/db\_xref="PID:d1039153"  
/db\_xref="PID:g4514433"  
/db\_xref="GI:4514433"  
/translation="DNSYVSGGLHGVGVSVYVNAALSKIELTVORAGQIHEDRYOGV  
POYPLRVGOTERTGTGKVPFAPAEFTSOTIFVNDILARLRSLFNLNGVRVLGDE  
RINLHEVDFEGGLSEFVYINCGKNHNLNIEFTTQAEKNGIGVEVALQWMSYQENV

RCFTNNIPKDGSTHLAGFRAALTGRNLNENILKKREKAVTGDAREGATLAVS  
VKYDPKSSOTKREKTVSSSEVRAVEONAKASVILENPAKAKITADITDAAR  
DAARAKEMTRRSALDIDIALPKLDCCKDPALSELFLVEDSDSGSKQGRNRKM  
QALPLKGLKILNERNRPFDMISSAEVGLITLALGGIGIEEYNPDKLRHKXIV"

BASE COUNT 317 a 230 c 302 g 318 t

Query Match 100.0%; Score 11; DB 1; Length 1167;  
Best Local Similarity 100.0%; Pred. No. 1.8e+03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTAGT 11  
|||||  
Db 940 CTGTTCTAGT 950

RESULT 4  
JG 1A61G/C

DEFINITION ACCAA61G 1154 bp DNA BCT 17-FEB-1994  
Acinetobacter haemolyticus aminoglycoside 6'-N-acetyltransferase  
(aac(6')-I<sub>g</sub>) gene, complete cds.

ACCESSION L09246  
NID 9455436  
VERSION L09246.1 GI:455436  
KEYWORDS aac(6')-I<sub>g</sub> gene; aminoglycoside 6'-N-acetyltransferase.  
SOURCE Acinetobacter haemolyticus (strain BM2685) DNA.  
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
Moraxellaceae; Acinetobacter.

REFERENCE 1 (bases 1 to 1154)  
AUTHORS Lambert, T., Gerbaud, G., Gallmand, M. and Courvaill, P.  
TITLE Characterization of the Acinetobacter haemolyticus aac(6')-I<sub>g</sub> gene  
encoding an aminoglycoside 6'-N-acetyltransferase which modifies  
amikacin

JOURNAL Antimicrob. Agents Chemother. 37, 2093-2100 (1993)  
MEDLINE 94079352

FEATURES  
Location/Qualifiers

1..1154  
/organism="Acinetobacter haemolyticus"  
/strain="BM2685"  
/db\_xref="taxon:29430"  
544..981  
/gene="aac(6')-I<sub>g</sub>"  
544..981  
/gene="aac(6')-I<sub>g</sub>"  
/codon\_start=1  
/transl\_table=11  
/product="aminoglycoside 6'-N-acetyltransferase"  
/protein\_id="AA21889.1"  
/db\_xref="PID:9455437"  
/db\_xref="GI:455437"  
/translation="MNKPASASLKDLELRNKMSPDSEASHLOEMHOLAETKALD  
LLAVSDHQAIALMASIRFEYNGCTSPVGLSEITVLPARRSGVATMLIRQAEVW  
AKQSCPEFASDAADNVTSHMRSLGFEETKVVYSKID"

BASE COUNT 322 a 200 c 260 g 372 t  
ORIGIN

Query Match 100.0%; Score 11; DB 1; Length 1154;  
Best Local Similarity 100.0%; Pred. No. 1.8e+03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTAGT 11  
|||||  
Db 432 CTGTTCTAGT 422

RESULT 5  
ACCGYB12 300 bp DNA BCT 04-FEB-1999  
LOCUS Acinetobacter sp. gene for DNA gyrase subunit B protein, C terminal  
side, partial cds.

ACCESSION D73423  
NID 91322104  
VERSION D73423.1 GI:1322104  
KEYWORDS DNA gyrase subunit B protein.  
SOURCE Acinetobacter sp. (strain:SEIP 12.81) DNA.  
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
Moraxellaceae; Acinetobacter.

REFERENCE 1 (bases 1 to 300)  
AUTHORS Yamamoto, S.  
TITLE Direct Submission  
JOURNAL Submitted (30-SEP-1995) to the DDBJ/EMBL/GenBank databases. Satoshi  
Yamamoto, Marine Biotechnology Institute, Kamaiishi Laboratories;  
3-75-1 Heita, Kamaiishi, Iwate 026, Japan  
(E-mail:HGPD02251enftfserve.or.jp, Tel:0193-26-6538,  
Fax:0193-26-6584)

REFERENCE 2 (bases 1 to 300)  
AUTHORS Yamamoto, S.

JOURNAL Unpublished (1996)  
REFERENCE 3 (sites)  
AUTHORS Yamamoto, S. and Harayama, S.  
TITLE Phylogenetic analysis of Acinetobacter strains based on the  
nucleotide sequences of gyrB genes and on the amino acid sequences  
of their products

JOURNAL Int. J. Syst. Bacteriol. 46 (2), 506-511 (1996)  
MEDLINE 97088999

FEATURES  
Location/Qualifiers

1..300  
/organism="Acinetobacter sp."  
/strain="SEIP 12.81"  
/db\_xref="taxon:472"  
/note="correspond to pos.1183-1482 in E.coli K-12 gyrB  
gene (X04341, X00870)."  
1..300  
/gene="gyrB"  
<1..>300  
/gene="gyrB"  
/transl\_table=1  
/codon\_start=1  
/product="DNA gyrase subunit B protein"  
/protein\_id="BA11148.1"  
/db\_xref="PID:d1011809"  
/db\_xref="PID:g1322105"  
/db\_xref="GI:1322105"  
/translation="KSAIDIDAGLPKRIADQCKDPALSELFLVEGDSAGSANOGRNR  
KMAIDLPKGLNVERARFDMISSAEVGLITLALCGIGIEEYNPDKLRHKIV"

gene

CDS

BASE COUNT 75 a 57 c 86 g 82 t  
ORIGIN

Query Match 100.0%; Score 11; DB 1; Length 300;  
Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTAGT 11  
|||||  
Db 76 CTGTTCTAGT 86

RESULT 6  
AF027868 87500 bp DNA BCT 15-NOV-1997  
LOCUS Bacillus subtilis chromosome region between terc and odnab.  
DEFINITION AF027868  
ACCESSION AF027868  
NID 92618993  
VERSION AF027868.1 GI:2618993

ORGANISM Bacillus subtilis.  
Bacteria; Firmicutes; Bacillus/Clostridium group;  
Bacillus/staphylococcus group; Bacillus.  
REFERENCE 1 (bases 59208 to 59540)  
AUTHORS Muller, J., Walter, F., van Dijk, J.M. and Behnke, D.

TITLE Suppression of the growth and export defects of an *Escherichia coli* secA(Ts) mutant by a gene cloned from *Bacillus subtilis*

JOURNAL Mol. Gen. Genet. 235 (1), 89-96 (1992)

MEDLINE 93062812

REFERENCE 2 (bases 34592 to 35233)

AUTHORS Wolf, M., Geczi, A., Simon, O. and Boriss, R.

TITLE Genes encoding xylan and beta-glucan hydrolyzing enzymes in *Bacillus subtilis*: characterization, mapping and construction of strains deficient in lichenase, cellulase and xylanase

JOURNAL Microbiology 141 (Pt 2), 281-290 (1995)

MEDLINE 95219081

REFERENCE 3 (bases 1 to 87500)

AUTHORS Lapidus, A., Galleron, N., Sorokin, A. and Ehrlich, D.

TITLE Sequence analysis of the *Bacillus subtilis* chromosome region between the *terc* and *odhAB* loci cloned in a yeast artificial chromosome

JOURNAL Unpublished

REFERENCE 4 (bases 1 to 87500)

AUTHORS Lapidus, A., Galleron, N., Sorokin, A. and Ehrlich, D.

TITLE Direct Submission

JOURNAL Submitted (01-OCT-1997) Laboratoire de Genetique Microbienne, INRA, Domaine de Vilvert, Jouy-en-Josas cedex 78352, France

FEATURES

source Location/Qualifiers

1. 87500

gene /organism="Bacillus subtilis"

db\_xref="taxon:1423"

1..560

gene="yoxB"

/gene="yoxB"

/citation={3}

1..560

gene="yoxB"

/citation={3}

/codon\_start=3

/transl\_table=11

/product="yoxB"

/protein\_id="AAB84418.1"

db\_xref="PID:92618994"

db\_xref="GI:2618994"

/translation="AYETAESLLGGTLEONAVDTKIAAGSGSMRNLITMDLGG LPQAFQISAAALFGLRMLHTTSVIXKPVPIQOKNYGXPAITHSPISTVAFGH FPAELNHYVGPALLIPLGKAETVCEITLRHSILONLCLNCFPPSGANGHRLKQFS KKEQLEORIRFEPALVDPAIEKRK"

604..1212

gene="yoaA"

604..1101

gene="yoaA"

/note="similar to B.subtilis p20 protein and E.coli ribosomal protein alanine acetyltransferase (194 aa)"

/citation={3}

/codon\_start=1

/transl\_table=11

/product="putative alanine acetyl transferase"

/protein\_id="AAB84419.1"

db\_xref="PID:92618995"

db\_xref="GI:2618995"

/translation="MEPIEDRLRLROITDODARAFACFSNDEVTRYGLNNESI EQAISMIOFALYQEKRGIRKRGIRKOTKIGITGFHALQKRRKEIGETIIEH WRNGFASEYISKVSYSYFSAUGLSRIGAVETIDNASNRLIKMGFORECVLROYMS KWDSI"

1187..1212

terminator /gene="yoaA"

complement(1217..2461)

gene /gene="yoaB"

complement(1217..2461)

CDS /gene="yoaB"

/note="similar to B.subtilis multidrug resistance protein 2 (400 aa)"

/citation={3}

/codon\_start=1

/transl\_table=11

/product="putative transporter"

/protein\_id="AAB84444.1"

db\_xref="PID:92619020"

db\_xref="GI:2619020"

/translation="MLDKIGIPKLANGLFGLVAFMGDGLFQGLSPFLNGLTVO OSASTFSTIGTALAIASWFSVCLEAFKAKTMTGTLFVIGTAAFTVFEDINLP VMAYTVFVKGLGYPFLAFSLFWIYRTPOSKRSTAVMGWIAVCLGFWGVAMYSY AIKAFVGLNTLMSSIFWVCLGAFALFINKDRPFKRRKRSRTAEILKGVYITLTPN RVLGGIRIRINSIGTYGFPVFLPMHMAHGISNVMQIIGTIFGLIVNLFGLYV GDRGKMTVFWPGGCGITVLLYTAAPVSGGSAVSVISGIFGGLGALVYVIGA IVPIVAGDKRAANSVNLNAGLSAFSPALAMFLIGLVAGQVWIFRALLYASAVL TKCIHPEKAKVEETSPQVNS"

complement(2555..4018)

gene /gene="yoaC"

complement(2555..4018)

/gene="yoaC"

/citation={3}

complement(2555..4018)

gene="yoaC"

/note="similar to Lactobacillus xylulose kinase (501 aa)"

/citation={3}

/codon\_start=1

/transl\_table=11

/product="xylulose kinase"

/protein\_id="AAB84445.1"

db\_xref="PID:92619021"

db\_xref="GI:2619021"

/translation="MKKQGYLFDIGTGNARVAVSVTSGSVLTVREDIRYETELT PDSRYFSPVAMKQVMTAKRALSQDIDITGLTSGRQGIYLIQNSPRTGLP INDRGRENAEIPMEETYSSTGRPLTALBALKILGKQKOPSLMERTKASTISDM VTQVLSGLTYEPQATEETLFDVQKNTSWSEKMDIFGFSILPLVRACTAGTGA NEVASSELGSLINAKVIAGGDTOLAVSTGSLDIVVSCSTPITITEDHGTGKA KAMLNCHTDGHLVETNPGITGNOKLQIFYPNETYEVNEEISALAKEDNACVA ALGSYLSERKALTRGGFLPDAPLAAHAKRAHFRALEETAFISKWEDITLTPFP ERDVMVCGGFGQSKALQYIADLQKRYVQDEYHQASVGAIVICNETQTLEENS ANVAVIEPKDQIELEAEEMKQTFRFGSGESKVL"

complement(4036..5070)

gene="yoaD"

complement(4036..5070)

/gene="yoaD"

/note="similar to B.subtilis D-3-phosphoglycerate dehydrogenase (419 aa) encoded by *serA* gene"

/citation={3}

/codon\_start=1

/transl\_table=11

/product="yoaD"

/protein\_id="AAB84446.1"

db\_xref="PID:92619022"

db\_xref="GI:2619022"

/translation="MKNTMKRMFCMTVLVYAPYNEGRKLENLFGSVAYQSKREG RAYREDELIQLKATNATGLTLEDQVDSVFAVPELSEFGVGRGMPNSVDVAASK RGIIVFTPGNNAQAVAMETIGNISLRHSASNMOKQEMSDVLOAVKRGNE LRGKVMICMGAGGARIKILITFAPDKIKKXYDPYIODDRLYKASIKTYFSSDIY SVHLPRTEITGLDROYFDKMSALFVNTSRAVVNRREDLTVLKEHKTSGLIADV FYHEPPESEDELISLPNVLATPRLAGATEVEEDHVTIINKALKKMGKXTLNIQTM YNKDALKTGG"

5394..7427

gene="yoaE"

5394..7427

gene="yoaE"

/note="similar to B.subtilis yae protein (667 aa) and E.coli formate dehydrogenase (715 aa)"

/citation={3}

/codon\_start=1

/transl\_table=11

/product="putative formate dehydrogenase"

/protein\_id="AAB84420.1"

db\_xref="PID:92618996"

db\_xref="GI:2618996"

/translation="MRSFATQONGIFKVCSLDPCDQGLIHKKQGIKYVQGDPDH PYTAGNICKNRYNMTERTIYDEKRTITPKRTGAKGAIPEPISMKKAITDSMKWL IDEBGAELIPYFYGNMGKTLRAGMRFRFFRSGSQQLERTYTSKAGSGCYRTWGI SAGIDPEETVHTKFLFWGINAVSTNHQITLQAKKAKKAKIIVDVHANNQGTSLND WFTPIKPGTDSALALGIMHILFKENLHDEAFLESYVGYELREHVAQYDPEKVSIT GVSVEDIYRLAKYGETSPSIFIRIGNPDHNDGNGMVARTIACLPATGOMLHAGGA IKHNSGILEENTNALQRPDLKGRTPSPFNNOGQVRLTDPRIELFTYGTNPAYV APEANKRQGLNREDLFTVYHDLFETLTAAYADIVLPATSAFENTDRTYSTWYHYIOL QOPIERYGSEKSTEVFRLLAEMGTTDELDKSDDEVILRQALDHPDNHLEAIDID

SLTKSEMAKREKPLFPGELPTPSGKILEYSEKMKODGFPALPTYPTVNDNEHPFM  
YVGPNNHNFLNTPSNNEKHLEKTPKIFITKDAEKGIYDGAIPVRWNRSCECL  
TAAVEQVLPGVVVSOGGLMADROGKKTACKRANPSLVNRGRRLFFRPCTD"

gene  
terminator  
complement(9379..11027)  
/gene="yohA"  
complement(9379..9406)  
/gene="yohA"  
complement(9423..11027)  
/gene="yohA"  
/note="similar to B. subtilis methyl-accepting chemotaxis  
protein McpA (661 aa)"  
/citation=[3]  
/codon\_start=1  
/transl\_table=11  
/product="yohA"  
/protein\_id="AAB84447.1"  
/db\_xref="pid:g2619023"  
/db\_xref="gi:2619023"  
/translation="MISSVKKNEELKDKMEROKETKHITQVELTGISNDEGLITRD  
KEIDBGMKADVDYKSDRVNDLIDEEYQSNIEDIKTSFYQIRALNDQVTAISSN  
PKKAEIHFGEERTLRKEGVAPAVNKLSDRLQVEEDLDEIQNGKMSQLITVG  
IYVILGIYLSIMLKSIWPLRSINKOLEIHEADLTAKVYVKNKDEFGQLAOSFN  
STHSITQIVKQISSSEQVAASEELASAESEKSTSHIRAMOMADSNVKSMS  
TEKSAESTRELDSTSSVANTGNITADSSMRDRAEIGSKVNMKLDQMKITKSDV  
SAGNLCITVASTAEISDSSITTSSEPTNLALNAATEARAGEOCGRVAAEV  
RKLADEINKSANHIOGVATTONESIEIYVNTKIVQENYSSGIVSQETGQNFELN  
LVEQVTSQIQEVAANQQLTSGEVNIVHTLAQKETSANTVAANSODELHSM  
GEISYAAESLSQIAELQTVINRFKY"  
11755..12910  
/gene="yohA"  
/citation=[3]  
11753..12882  
/note="yohA"  
/gene="yohA"  
/note="similar to H. influenzae HI0035 protein (551 aa)"  
/citation=[3]  
/codon\_start=1  
/transl\_table=11  
/product="transcription antiterminalator"  
/protein\_id="AAB84421.1"  
/db\_xref="pid:g2618997"  
/db\_xref="gi:2618997"  
/translation="NNTVMVSFASSAELLKDKENCPPEHILMDYBQAKHDSFTHTF"

Query Match 100.0%; Score 11; DB 1; Length 87500;  
Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CTTCTTCTACT 11  
|||||  
Db 65049 CTGTTCTACT 65039

RESULT 7  
LOCUS BACS18422/c 2985 bp DNA BCT 23-NOV-1998  
DEFINITION Bacillus sporothermodurans partial uvrB and uvrA genes.  
ACCESSION Y18422  
NID 93924674  
VERSION Y18422.1 GI:3924674  
KEYWORDS uvra gene; uvrb gene.  
SOURCE Bacillus sporothermodurans.  
ORGANISM Bacillus sporothermodurans.  
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;  
Bacillus.  
1 (bases 1 to 2985)  
REFERENCE  
AUTHORS Herman, L. and Heyndrickx, M.  
TITLE Bacterial typing by REP-PCR can be based on the presence of random  
homologous sequences  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2985)  
AUTHORS Herman, L.M.F.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-1998) L.M.F. Herman, Centre of Agricultural

Research, Department Animal Product Quality, Brusselssesteenweg 370,  
9090 Melle, BELGIUM

FEATURES  
source  
Location/Qualifiers  
1..2985  
/organism="Bacillus sporothermodurans"  
/strain="MB342"  
/db\_xref="taxon:46224"  
<1..1784  
/gene="uvrB"  
/product="uvrB protein"  
1..1784  
/gene="uvrB"  
1792..>2985  
/gene="uvrA"  
/product="uvrA protein"  
1792..2985  
/gene="uvrA"

gene  
BASE COUNT 994 a 497 c 655 g 831 t 8 others

ORIGIN

Query Match 100.0%; Score 11; DB 1; Length 2985;  
Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CTTCTTCTACT 11  
|||||  
Db 2467 CTGTTCTACT 2457

RESULT 8  
LOCUS BARRPEG/c 4753 bp DNA BCT 24-JUN-1994  
DEFINITION B. aphidicola trpE and trpG genes for anthranilate synthase,  
components I and II.  
ACCESSION Z21938  
NID 5472880  
VERSION Z21938.1 GI:472880  
KEYWORDS anthranilate synthase I; anthranilate synthase II; trpE gene; trpG  
gene.  
SOURCE Buchnera aphidicola.  
ORGANISM Buchnera aphidicola.  
Bacteria; Proteobacteria; gamma subdivision; Buchnera.  
1 (bases 1 to 4753)  
REFERENCE  
AUTHORS Lai, C.Y., Baumann, L. and Baumann, P.  
TITLE Amplification of trpEG: adaptation of Buchnera aphidicola to an  
endosymbiotic association with aphids  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (9), 3819-3823 (1994)  
MEDLINE 94224833  
REFERENCE  
AUTHORS Lai, C.  
TITLE Direct Submission  
JOURNAL Submitted (10-MAR-1993) Chi-Yung Lai, Microbiology, University of  
California-Davis, Davis, California, 95616-8665, USA  
FEATURES  
source  
Location/Qualifiers  
1..4753  
/organism="Buchnera aphidicola"  
/db\_xref="taxon:9"  
1..1728  
/note="this 1.7kb EcoRV fragment is cloned separately and  
not necessary contiguous with the rest of the sequence."  
1..3580  
/note="One unit of a four-unit plasmid"  
evidence-experimental  
1027..4753  
/gene="trpE"  
1027..2574  
/gene="trpE"  
/codon\_start=1  
/transl\_table=11  
/product="anthranilate synthase, component I"  
/protein\_id="CAA79931.1"

```

/db_xref="PID:q472881"
/db_xref="GI:472881"
/db_xref="SWISS-PROT:P42387"
/translation="MSKNPYEIEIIQKTAHYPPDPTMAFNHLCASRPGTLLLETAENV
KRDESIIMIDSAMNISESDSVKLTPLSINGTILSTLTKTIKPIKLEIYKNNSTI
LVEPKIKNIDEDKLFSLVDFAPRLMIRIENNEKSKAFPGGLFSYDLISVEES
LPRKGNKNSCNECYLAETLLVLDHOKTKCLIONSLEPSKNEKERRIKKRSEVIER
LNEKLIKPKTKIKOINLNSNMNFEYCITIKLOKLIKOPVSRKRYLDCPN
PSAYOKLKNSPSPMFPMQODTTLFGASESSSKIDEKTRKLELPPIAGTRRGK
TGDNDLDDLSRIEMKTNKELAEHLMLVDLARNDLARICKGSRSYSDLVADR
YSHVHLVLSRVIGELNEGDLAAHAYASCMNMGTLGAPVRAMOLIAEHEGERSYSG
GAIGYFTDGNLDTCTTIRSAVYKQAVITQAGIVYNSIPENVESLNQAQAVIN
AIKNAY"
2592..3182
/gene="trpE"
/codon_start=1
/transl_table=11
/product="anthranilate synthase, component II"
/protein_id="CAA79932.1"
/db_xref="PID:q472882"
/db_xref="GI:472882"
/db_xref="SWISS-PROT:P42388"
/translation="MANILLDNFDSFTYNNLVEQLRNKNSVLIYRNTVDINTLINSI
KRNPILMLSPGSPSTPKNAGCMNLIRKKEIPVIGLGHQIVAYAGIITVAG
EIFHGKASLIMHGLEMEFGLPQPLPVARYSHLICKNIPKPNILINSYNDMIMSVRN
LDVCGFOPHPSILITSGALLLEKILIMASIKYKG"
4605..>4753
/gene="trpE"
/codon_start=1
/transl_table=11
/product="anthranilate synthase, component I"
/protein_id="CAA79933.1"
/db_xref="PID:q472883"
/db_xref="GI:472883"
/db_xref="SWISS-PROT:P42387"
/translation="MSKNPYEIEIIQKTAHYPPDPTMAFNHLCASRPGTLLLETAENV
KRRL"
BASE COUNT      1949 a      615 c      599 g      1590 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 11; DB 1; Length 4753;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 CCGTTCTAGT 11
      1111111111
      4725 CCGTTCTAGT 4715
RESULT 9
LOCUS      BS16S138      970 bp      mRNA      BCT      08-SEP-1997
DEFINITION      Bacterium sp. 16S ribosomal RNA, clone 400-138.
ACCESSION      AJ001424
VERSION      AJ001424.1 GI:2385346
KEYWORDS      16S ribosomal RNA; 16S rRNA.
SOURCE      unidentified bacterium.
ORGANISM      Bacteria; environmental samples.
REFERENCE      1 (bases 1 to 970)
AUTHORS      Zaychikov E.F.
TITLE      Direct Submission
JOURNAL      Submitted (07-SEP-1997) Zaychikov E.F., Laboratory of Molecular
Enzymology, Limnological Institute, Ulan-Batorskaya str. 3,
Irkutsk, 664033, RUSSIA
REFERENCE      2 (bases 1 to 970)
AUTHORS      Denissova L.Y., Belkova N.L., Tulokhonov I.I., Parfenova V.V. and
Zaychikov E.F.
TITLE      Phylogenetic analysis of aquatic bacteria of the lake Balkal
JOURNAL      Unpublished
FEATURES      Location/Qualifiers

```

```

SOURCE      1..970
            /organism="unidentified bacterium"
            /db_xref="taxon:2338"
            /clone="400-138"
            <1..>970
            /gene="16S rRNA"
            /product="16S ribosomal RNA"
            /gene="16S rRNA"
            /gene="16S rRNA"
BASE COUNT      229 a      233 c      311 g      196 t      1 others
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 11; DB 1; Length 970;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 CCGTTCTAGT 11
      1111111111
      575 CCGTTCTAGT 585
RESULT 10
LOCUS      BSUB0003/c      209100 bp      DNA      BCT      26-NOV-1997
DEFINITION      Bacillus subtilis complete genome (section 3 of 21): from 402751 to
611850.
ACCESSION      299106 AL009126
VERSION      92632653
SOURCE      299106.1 GI:2632653
ORGANISM      Bacillus subtilis.
            Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
            Bacillus.
REFERENCE      1 (bases 1 to 209100)
AUTHORS      Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
Azevedo V., Berteiro M.G., Bessières P., Bilotin A., Borchert S.,
Borriss R., Boursier L., Brans A., Braun M., Brignell S.C.,
Bron S., Brouillet S., Brusch C.V., Caldwell B., Capuano V.,
Carter N.M., Choi S.K., Codani J.J., Connerton I.F., Cummings N.J.,
Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehlich S.D.,
Emmerson P.T., Entlan K.D., Errington J., Fabel C., Ferrari E.,
Folger D., Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A.,
Galleron N., Ghim S.Y., Glaser P., Goffeau A., Golightly E.J.,
Grandi G., Guisepi G., Guy B.J., Haga K., Hatach J., Harwood C.R.,
Haupt A., Hilbert H., Holsappel S., Hosono S., Hullo M.F.,
Itaya M., Jones L., Joris B., Karamata D., Kasahara Y.,
Klaerr-Blanchard M., Klein C., Kobayashi Y., Koetter P.,
Koningstein G., Krogh S., Kumano M., Kurita K., Lapidus A.,
Lardinois S., Lauber D., Lazarevic V., Lee S.M., Levine A., Liu H.,
Masuda S., Maueil D., Medigue C., Medina N., Meliello R.P.,
Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M.,
Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M.,
Portetle D., Porwollik S., Prescott A.M., Presecan E., Pujic P.,
Punelle B., Rapoport G., Ray M., Reynolds S., Rieger M.,
Rivola C., Rocha E., Roche B., Rose M., Sadale Y., Sato T.,
Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J.,
Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
Takeuchi M., Tanakoshi A., Tanaka T., Terresta P., Tognoni A.,
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
Viari A., Wambolt R., Wedler E., Wedler K., Weltergasser T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H. and Danchin A.
TITLE      The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
JOURNAL      Nature 390 (6657), 249-256 (1997)
MEDLINE      98044033
REFERENCE      2 (bases 1 to 209100)
AUTHORS      Kunst F., Ogasawara N., Yoshikawa H. and Danchin A.
TITLE      Direct Submission

```



## JOURNAL

Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,  
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724  
Paris Cedex 15, FRANCE. E-mail: mosze@pasteur.fr,  
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45  
68 89 48

## FEATURES

```

source      Location/Qualifiers
            1..209100
            /organism="Bacillus subtilis"
            /strain="168"
            /db_xref="taxon:1423"
            48..1304
            /gene="ycxa"
            48..1277
            /gene="ycxa"
            /function="unknown"
            /codon_start=1
            /transl_table=11
            /protein_id="CAB12161.1"
            /db_xref="PID:e1182320"
            /db_xref="PID:g2632654"
            /db_xref="GI:2632654"
            /db_xref="SWISS-PROT:008791"
            /translation="MKTSRSKRVLEFFVPAIGNSPNSFOEFLMPADFAADR
            SLISVSIFMITTGIOPEFEDRSVKIMAGVCSASFVLPSPNVAHS
            AIVGIGIGYCAVGTVOYFISCMEDHKGILALILNANSAGLVSPPIAAR
            VYAGWOSTYITIGYMAVLPVLPFGMKRPHQAQETKSYDWRGNVAKQSLI
            HILYGVFCGTGIGIDNAHLVPLINKDAVSVNMMAGFAFIIGGLAMLDL
            GSRVMSLTFEIRLSLCLIPILGISHDLMTFGILFGLSYGVPIPTASIS
            ESYQTLIGSLGILNFIHQVAGLSVAGLGFPMTHGYLILVAVCIYFVGLSAIV
            LVPFLDKOKAKETHSI"
            1280..1304
            /gene="ycxa"
            complement(1292..1849)
            /gene="ycxb"
            complement(1292..1849)
            /gene="ycxb"
            /function="unknown"
            /codon_start=1
            /transl_table=11
            /protein_id="CAB12162.1"
            /db_xref="PID:e1182321"
            /db_xref="PID:g2632655"
            /db_xref="GI:2632655"
            /db_xref="SWISS-PROT:008793"
            /translation="MIOYASESINLPGETFEKDVREIFFYQIAKISCYFLLCAIFA
            AVNFINGPRIYVGSDALFNNSMLITVSYLFTLLLLLYVFSRAKKNEKMS
            KRTYLNQEGIRYICSKKYDLFNMDETAVFEYKNIFRVNTSSGOYIAPKHFSESE
            EMNFEKILIKNETEKRLKFKKDH"
            complement(1903..2841)
            /gene="ycxc"
            complement(1903..2841)
            /gene="ycxc"
            /function="unknown"
            /note="similar to hypothetical proteins from B. subtilis"
            /codon_start=1
            /transl_table=11
            /protein_id="CAB12163.1"
            /db_xref="PID:e1182322"
            /db_xref="PID:g2632656"
            /db_xref="GI:2632656"
            /db_xref="SWISS-PROT:008794"
            /translation="MYKHOQTPATYATAILYSPITLSFLVKIAIQTAEPEDILAHFR
            TIAFAATVPILFGVWKLILKVDILPLALPLAFESFOAGLVYSSSESGIIL
            QAIPFTWVAAYLVKERPTWKGFTLVSVAGWTFVYKNGVDVESAKGSLIL
            LSALSAMYNARAKMTQRFKTELTLYLSAIGVFNALVHRGAGAGTGYEFLP
            RPPGVLAIVGLSVLSTFSLSNYTSRIEAFKGSAPFNHSTIVTIAGVILNES
            LAMVHLAGAVCMIGVGSNINLEKKTKRPGPAK"
            2965..4299
            /gene="ycxd"
            2965..4299
            /gene="ycxd"
            /function="unknown"
            /note="similar to transcriptional regulator (GntR family)

```

```

            /aminotransferase (Mocr-1like)"
            /codon_start=1
            /transl_table=11
            /protein_id="CAB12164.1"
            /db_xref="PID:e1182323"
            /db_xref="PID:g2632657"
            /db_xref="GI:2632657"
            /db_xref="SWISS-PROT:008792"
            /translation="MEKYSLSLIRIEMQSTVQEGDRLPSIROLARSYOVSKSTVI
            RALQELERKHLIYSVKSGYIVKSKGSKSGOPIDFATSPADVPYLDPOHCI
            NKAIDYKNDLFTYGTGPKLPSLIVLRKRLATQOVFDERHIFITSGVOQLSLCA
            MPFGKREKIAIEQGYHLMWOLETLEIPAIQYKRETEGDIATKVEPLFOESTKFA
            YTMPRPHPLGCSLSEKROKELVRAEYVDIYVEDDLGDEENKRAKDPATYDLS
            HYTLKSTSKMFPGLRVGAALVPALTDGTGYAKKLNDDICSMISQALAEIYKSGM
            YGRHEKIRDSYKERSRLHQAIRTRDOLGSGRFTFSSGOAPCMHTHLVLPDLPASR
            VIHRLKQGVLEAIDRHYLSDPENKLKINISVAKTEIDIERGVKILMSHL"
            complement(4472..4969)
            /gene="sfp"
            complement(4472..4969)
            /gene="sfp"
            /codon_start=1
            /transl_table=11
            /product="surfactin"
            /protein_id="CAB12165.1"
            /db_xref="PID:e1182324"
            /db_xref="PID:g2632658"
            /db_xref="GI:2632658"
            /translation="MKIYGIANDRPLSOENRPMFTISPKREKRCRRFYKEDAHRT
            ILGDVLSVLSROYOLDKSDIRSTQBYGKPCYDLPDPAFNTSHSGRWYIGAFDSO
            PGIDIEKTRPISLEIARFPFSKTEISDLKDKDEQDITYTHLSMKESPIKQGROR
            LKAS"
            complement(5053..5071)
            /gene="ycze"
            complement(5053..5721)
            /gene="ycze"
            complement(5074..5721)
            /gene="ycze"
            /function="unknown"
            /note="similar to hypothetical proteins from B. subtilis"
            /codon_start=1
            /transl_table=11
            /protein_id="CAB12166.1"
            /db_xref="PID:e1182325"
            /db_xref="PID:g2632659"
            /db_xref="GI:2632659"
            /db_xref="SWISS-PROT:024927"
            /translation="MKOELVIRMTFFVAGLIIAFVSTIEGRALGISPDADHYSL
            FOHGLYVGSIIITGALIVFTSLFTRAPKIGALNMLVIGVIFDEFTLPAPST
            YTGSIIVFSLGVILIGVGVYVAGAGGRDLSMLITEKTMNVQWANGMEITP
            LFAWGMGGPIGFQGITLAILTGLILFSLPOSTQLNLYAVARRTAVKASPVH"
            complement(5815..6782)
            /gene="ycxi"
            complement(5815..5830)
            /gene="ycxi"
            complement(6039..6782)
            /gene="ycxi"
            /function="unknown"
            /note="similar to glutamine ABC transporter (ATP-binding
            protein)"
            /codon_start=1
            /transl_table=11
            /protein_id="CAB12167.1"
            /db_xref="PID:e1182326"
            /db_xref="PID:g2632660"
            /db_xref="GI:2632660"
            /db_xref="SWISS-PROT:P39456"
            /translation="WLVKGLNKSFGENILUKIDMKIEKGVYAILIGPSGSGKTTLL
            RCLNALIEIPRNGELAFDDFSIDSKVKVQADIKLRKSPVYQVAHLPFRALENV
            MEGPVQVQKRKKEVERKEAIOLDKVGILDKMDLPPQLSGGQOQVGIARALAIQPE
            LMDFDEPTSLADPELVGEVLKVIDLANEGMTWVVYHEIKFAQVADDEVIFIDGVYI
            VEGQPPQIFSAPEERTFORILNLNPL"
            complement(6796..7500)
            /gene="ycxj"

```





complement(99. .1001)  
/gene="ytm"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB14906.1"  
/db\_xref="pid:e185801"  
/db\_xref="pid:g2635412"  
/db\_xref="gi:2635412"  
/db\_xref="SPTREMBL:O34430"  
/translation="MKRLTFAFIFGLISLQIDSLGNAYGVTSILLAFGIPPAVSAVHALEYVTASGVASHIKFGNDKQVVLIVDLSGALIGAAFLISQLPDAKRYISLFIILLAGVYLIRFLQYKPALEKHAVPLNRKOSILVGIAGFAOTGGGEPYTPILSRGLSGSRKAVGVTDSEFAIVASAAAGFLISGWDVAMLVWFSLMAGIIRAPIAMVLVOKPEPGLMAGVGVGFIILVAVRTLINEMIANTAVHPLIYTAIGAILVSALVLSLKISGRNRYKTAADVHLKEK"  
complement(1020. .2270)  
/gene="hnp0"  
/element(1020. .2270)  
/gene="hnp0"  
/EC\_number="3.5.1.32"  
/codon\_start=1  
/transl\_table=11  
/product="hnpurate hydrolase"  
/protein\_id="CAB14907.1"  
/db\_xref="pid:e185802"  
/db\_xref="pid:g2635413"  
/db\_xref="gi:2635413"  
/db\_xref="SPTREMBL:O34980"  
/translation="MSLDYMRNIGSSVYQTGNDILTKESNPVLSLLEKOLISIRRHLYQELSKKEEFETTAFTKCKKEKGIQIRFTALKEGVFADIAESESPPALAAADIDALPIEKGGLPSPASHKKGIMHACGHEFTTALLGAFLKGNOSLKGITLLEQPAEAGAGATVIEGDQGDIDAYIGLNKEDIVGVGLTGPLMAADVRFKVEIGKGAHALAHNENFDPIIGASOLIVALTQIIVSNVPLQSAIITVKGINGSWANYIPGTVVIETKATVPESEVNRVOKRFPAYTEBQISAFSLKKNVMSGPPCLNDEATGIVRDAKATLQVTDIPASPAGDEPAVYLEHITGSAFTFGTGDGDDHMHHPATIDETALIKASYRLVTSARLRLDNSEKISD"  
complement(2288. .2980)  
/gene="ribR"  
/element(2288. .2980)  
/gene="ribR"  
complement(2288. .2980)  
/gene="ribR"  
/function="regulation of riboflavin biosynthesis genes"  
/note="alternate gene name: ytmk"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator"  
/protein\_id="CAB14908.1"  
/db\_xref="pid:e185803"  
/db\_xref="pid:g2635414"  
/db\_xref="pid:g2635414"  
/db\_xref="gi:2635414"  
/db\_xref="SWISS-PROT:P94465"  
/translation="MTIAGTVVKGKOLGRKLPPTANVDAKINGLNRYGVYAAATVNHOFHILGVANNICVAKPVGNSNLEKTELEIFLFDPHRIYGEKITECSILFIRBERPDSLESLTRKIKKIDICVAKRPELTIGIAPNKNKESLSHOELNPLDLCYKKNLNDYGNRVINVDNMFEGEGITQVAYRIRIYILSFSLKEDNPVSKSIRFAGGLADKINRFLSYVESEENIIG"  
complement(3027. .4355)  
/gene="ytmJ"  
complement(3027. .4355)  
/gene="ytmJ"  
/function="unknown"  
/note="similar to ntrIIotriacetate monooxygenase"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB14909.1"  
/db\_xref="pid:e185804"  
/db\_xref="pid:g2635415"  
/db\_xref="pid:g2635415"  
/db\_xref="gi:2635415"  
/db\_xref="SPTREMBL:O34974"  
/translation="MTRADPIQFGMNLHGCGTTDGMKRRHPVDPSASTNIEFYKKRAQTAERKLSEFIRIADGLFISEKSIPIHLKRFEPITILSLASVYKNIGLVQTFSTSEFLPTTSRQIAMSDDHYSISGSGAGNNLVTSPQEGAAHRHSSSNLPEHTEREYLQEHLDVVRGLTMSWSEDAIDHNKKTGQFNDQAKLHRLNLNKGKGYFQVEGPLNIGRSKQCEPYPVQFG

```

gene
CDS
    complement(4352, .4633)
    /gene="ytn1"
    /function="unknown"
    /codon_start=1
    /transl_table=11
    /protein_id="CAB14910.1"
    /db_xref="PID:e1185805"
    /db_xref="PID:g2635416"
    /db_xref="GI:2635416"
    /db_xref="SPTREMBL:O34639"
    /translation="MSDVNIIVVMSKGCSCYCEVYVLEKGFPPONIDYSEKUR
DIIQVKGVRHVPVVEIGRGNQYGTGTEIGIEIHDILANHAQIKKAKR"
    complement(4648, .5652)
    /gene="ytm0"
    complement(4648, .5652)
    /gene="ytm0"
    /function="unknown"
    /note="similar to hypothetical proteins"
    /codon_start=1
    /transl_table=11
    /protein_id="CAB14911.1"
    /db_xref="PID:e1185806"
    /db_xref="PID:g2635417"
    /db_xref="GI:2635417"
    /db_xref="SPTREMBL:O34846"
    /translation="MIRLSIDQSLIGEGTADDTLQHTVKAQMAEECGYHRFWAE
HHNNDIAGSAPSEVLGLYLAESTPKIRILASGVYALQHSYKVAEOPHILSLAPGRI
DLGKAGAGGROSTDLQALQAEKVPKRPQPEKLEELTFVRDDEPTDRAALAPROO
YDRKPGIFLGGSTESHSIAKAGISVFVYFFINGEEVYLKERRATDAHLPROSEAE
PHLAPFAAHTKEAEKHLVSRSEISFVVLKDKKVNWSGEQAEVLENTVEYDII
YQKGIAGTKEEVAEELTRLSGTYKINDEYFTPIKNAVEKOLSYOLSDVAYLAARR
"
    complement(5649, .6428)
    /gene="hisp"
    complement(5649, .6428)
    /gene="hisp"
    /codon_start=1
    /transl_table=11
    /product="histidine transport protein (ATP-binding
protein)"
    /protein_id="CAB14912.1"
    /db_xref="PID:e1185807"
    /db_xref="PID:g2635418"
    /db_xref="GI:2635418"
    /db_xref="SPTREMBL:O34900"
    /translation="MIRKINIKQFGIHHVLYKGINLTVRKGEVYTIIGSGSGKTFLL
RCJLNLKRPDEGITISINDKVINCFRPSKKEVHMLRKQTAVFQOYHIFAKTVIENVM
EGGITAKMRKQDAIVSAENELKRVGQDLNAYPSQLSGGQKRVGIALALAHDPVY
LRLDEPRAALDELVEGVELMEIVATGATVIVTHEMFARVSDVQVMEGVIV
EOSTPEVFRHKTKEKRLROFLRRVSPYLEPEKHIKEPYI"
    complement(6425, .7132)
    /gene="ytmn"
    complement(6425, .7132)
    /gene="ytmn"
    /function="unknown"
    /note="similar to amino acid ABC transporter (permease)"
    /codon_start=1
    /transl_table=11

Query Match      100.0%; Score 11; DB 1; Length 215640;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY      1 CTGTTCTAGT 11
          |||||
Db 189460 CTGTTCTAGT 189450

```

```

RESULT 13
BTCLORGEN/C
LOCUS      1707 bp      DNA      BCT      25-JUL-1996
DEFINITION B.thuringiensis p1cr gene and orf2.
ACCESSION  X93361
VERSION     X93361.1 GI:1212804
KEYWORDS    ORF2; p1cr gene; p1cr protein.
SOURCE      Bacillus thuringiensis.
ORGANISM    Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
            Bacillus.
REFERENCE   1 (bases 1 to 1707)
AUTHORS     Lereclus,D., Agaisse,H., Gontnet,M., Salameau,S. and Sanchez,V.
TITLE       Identification of a Bacillus thuringiensis gene that positively
            regulates transcription of the phosphatidylinositol-specific
            phospholipase C gene at the onset of the stationary phase
JOURNAL     J. Bacteriol. 178 (10), 2749-2756 (1996)
EDLINE      96212013
REFERENCE   2 (bases 1 to 1707)
AUTHORS     Lereclus,D.
TITLE       Direct Submission
JOURNAL     Submitted (22-NOV-1995) D. Lereclus, Unite de Biochimie
            Microbiene, Dept des Biotechnologies, Institut Pasteur, 28 Rue du
            Dr. Roux, 75724 Paris Cedex 15, FRANCE
REFERENCE   3 (bases 1 to 1707)
AUTHORS     Lereclus,D.
TITLE       Direct Submission
JOURNAL     Submitted (01-MAR-1996) D. Lereclus, Unite de Biochimie
            Microbiene, Dept des Biotechnologies, Institut Pasteur, 28 Rue du
            Dr. Roux, 75724 Paris Cedex 15, FRANCE

FEATURES
    source
        1..1707
            /organism="Bacillus thuringiensis"
            /strain="407"
            /db_xref="taxon:1428"
            418..467
            /gene="p1cr"
            515..1372
            /gene="p1cr"
            515..1372
            /gene="p1cr"
            /function="positively regulates transcription of
            phosphatidylinositol-specific phospholipase C gene"
            /codon_start=1
            /transl_table=11
            /product="p1cr protein"
            /protein_id="CA63744.1"
            /db_xref="PID:e214719"
            /db_xref="PID:g1212805"
            /db_xref="GI:1212805"
            /db_xref="SPRMBL:045782"
            /translation="MQAKLGSPIKIRIVRGITOKOLSENICHOSEVSRISGAVNP
            SMDLGIANKLQIPITHFREVLIYSIDIEKKQFKQVIMLCOKRKIKYKVMNEL
            KKEVYPERQQLQWQYVAAVYALKVDYECLEIKLNDQLTGIDYONLYINA
            IANIVANGYKQKIDGLEFOILKQLALHNEFDVKNVNHAKALYLDISREESLYO
            VNKAIEISCRINSMALIGOLYVQRGECRLREVEAEIDVAYKKAFFDILEMAYK
            EALVNIKISR"
            1461..1607
            /gene="orf2"
            /gene="orf2"
            1461..1607
            /gene="orf2"
            /codon_start=1
            /transl_table=11
            /protein_id="CA63745.1"
            /db_xref="PID:e224565"
            /db_xref="PID:g1212806"
            /db_xref="GI:1212806"
            /db_xref="SPRMBL:045783"
            /translation="MKKLLIGSLTLTAMAMGSLGDTAFPKSIIISHNDQEVVAADL
            PPEF"

BASE COUNT      614 a      203 c      321 g      569 t

ORIGIN

```

```

Query Match      100.0%; Score 11; DB 1; Length 1707;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTGTTCTACT 11
Db      1676 CTGTTCTACT 1666

RESULT 14
BUHRDDG/C
LOCUS      4877 bp      DNA      BCT      22-JUN-1994
DEFINITION Buchnera aphidicola ribosomal inner membrane protein
            (rnpA-rmpH-dnaA-dnaN-gyrB) gene, complete cds.
ACCESSION  M80817
VERSION     M80817.1 GI:144144
KEYWORDS    dnaA gene; dnaN gene; gyrB gene; inner membrane protein; ribosomal
            protein; rnpA gene; rmpH gene.
SOURCE      Buchnera aphidicola DNA.
ORGANISM    Bacteria; Proteobacteria; gamma subdivision; Buchnera.
REFERENCE   1 (bases 1 to 4877)
AUTHORS     Lai,C.-Y. and Baumann,P.
TITLE       Genetic analysis of an aphid endosymbiont DNA fragment homologous to
            the rnpA-rmpH-dnaA-dnaN-gyrB region of Eubacteria
JOURNAL     Gene 113, 175-181 (1992)
EDLINE      92241666
FEATURES
    source
        1..4877
            /organism="Buchnera aphidicola"
            /db_xref="taxon:9"
            complement(<1..821)
            /note="60 kDa"
            /codon_start=1
            /transl_table=11
            /product="inner membrane protein"
            /protein_id="AA73145.1"
            /db_xref="PID:g551760"
            /db_xref="GI:351760"
            /translation="MELQNFPIFAFLVFSFLWQAMOSQSLKRNKNEENSPFHLN
            HKREKNOIILKNDVLRIVNMGDIEASLDYKQLNSPENKLEFSDFYQA
            OSLGKDGDLNSIKRIRIYSAKKFELGHELELVPIFIKNGITIKNLIK
            RGVAVVEYDINNINNKREIKENITGQAKQYKPKPKNDIYSGNALTFFGAASS
            DAKYKIKFDNIANHEHLHITTHGWMALQOYFVAVWPDINISSNITITSTL"
            complement(824..1081)
            /note="10 kDa"
            /codon_start=1
            /transl_table=11
            /product="inner membrane protein"
            /protein_id="AA73146.1"
            /db_xref="PID:g144146"
            /db_xref="GI:144146"
            /translation="MVKSTIVFCLTFEVSIVONYISFMPNSNCRFYPTOSTWILS
            LRKEVIGIITLIRLFRCHPLHGGSDDIVPLKIKDSEY"
            complement(1060..1404)
            /gene="rnpA"
            /gene="rnpA"
            complement(1060..1404)
            /gene="rnpA"
            /codon_start=1
            /transl_table=11
            /protein_id="AA73147.1"
            /db_xref="PID:g144147"
            /db_xref="GI:144147"
            /translation="MNVNFKKSKSLKSTNQVYFSNPCKNFEHIIILGRSNLGH
            PLGSLSRKNIKHAYRNKIKRLIRETFPLQLHRLIMDFVYIAKKNIYVLLNNKIIV
            NILEYIWSNOR"
            complement(1418..1561)
            /gene="rmpH"
            /gene="rmpH"
            complement(1418..1561)
            /gene="rmpH"

```

```

/codon_start-1
/transl_table-1
/product="ribosomal protein"
/protein_id="AA073148.1"
/db_xref="PDB:9144148"
/transl_table="MKTFFPSILKRNHSGFRTMATKNGRIILSRRAKLRTLY
SSK"
2049..3344
/gene="dnaa"
2049..3344
/gene="dnaa"
/codon_start-1
/transl_table-1
/protein_id="AA073149.1"
/db_xref="PDB:9144149"
/transl_table="MWTSLKAKLNINLEIYAPNOFILDWKDKYLHPRKILQDFC
GNSPFIKYYOTSKERKNILOKONLAPIDKIPFESSHRSINNKHSF
ENFLEKSNOLARAASOVAKNPNSSNPLEYGTGKTHLLAINGLIAKXAV
KIYHSERFYODMVKALONNAIEKFLYYSVALLLIDIOFPAHKEBSOEFFHTF
NALLEGNOIILISDRPKELNGVEDLKSIFGGLTVAIDPELETRVALILKAE
NIVLSEVAFIAKHLNSVRELEGLNRYIVANFTHSITVEFVREALDLALQ
EKLYTIANIOKTVAEYKIKVADLSRRSRVAPROMAMAKELTNHSLPEIGDA
FSGDHTVTLACRKIEQLRKEMHDIKEDFNLIRTLISV"
3349..4449
/gene="dnaa"
3349..4449
/gene="dnaa"
/codon_start-1
/transl_table-1
/protein_id="AA073150.1"
/db_xref="PDB:9144150"
/transl_table="MFTIIONDIJTNLKKITRYLVNINISFPILENTLIOVEDTSL
TTNLEILSKIEIITRYIKGKTTISGRKILICRTLESEKIKMOLKKMYISE
NSNYILSLADTEPNHONFYSKIPDISNLIKEMTEKESKODVAYLNGML
EKKDFSVATDGYRLAISTOLKDNFSSITPKAYMELIKLNTOPOLNITI
GNSIRITKTLITTOIIBGEYPIYKSVLEKKNPITINSILKSLVLAHLAE
KFCGIEIKENGKRVLSDNOEETADLEIDYFGKIEISINWYLLDVINNISE
NIALFLNKSIOIEANNSNAYVMLKR"
4604..4877
/gene="gyrB"
4604..4877
/gene="gyrB"
/partial
/gene="gyrB"
/codon_start-1
/transl_table-1
/protein_id="AA073151.1"
/db_xref="PDB:9551761"
/db_xref="GI:551761"
/transl_table="MIDYDSKIKILRGDAVAKRPGMYIGDTDGSGLAHNVETIY
DNSIDELALGFCFEIKVYIHSDNSVSKDGDGRIPTDIHPEKISMA"
BASE COUNT 1955 a 617 c 605 g 1700 t
ORIGIN

Query Match 100.0%; Score 11; DB 1; Length 4877;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGTTCTAGT 11
|||||
DB 1434 CTGTTCTAGT 1424

```

```

RESULT 15
BUHTRPEA/c 3641 bp DNA BCT 14-AUG-1996
LOCUS Buchnera aphidicola anthranilate synthase large subunit (trpG) gene
DEFINITION and anthranilate synthase small subunit (trpG) gene, complete cds.
ACCESSION L43551
G1082002

```

```

VERSION L43551.1 GI:1082002
KEYWORDS anthranilate synthase; large subunit; small subunit.
SOURCE Buchnera aphidicola DNA.
ORGANISM Buchnera aphidicola
REFERENCE 1 (bases 1 to 3641)
AUTHORS Rouhachsh.D., Lai.C.Y., von Dohlen.C.D., Clark.M.A., Baumann.L.,
Baumann.P., Moran.N.A. and Voegelin.D.J.
TITLE The tryptophan biosynthetic pathway of aphid endosymbionts
JOURNAL (Buchnera): genetics and evolution of plasmid-associated
MEDLINE anthranilate synthase (trpG) within the aphididae
FEATURES
source J. Mol. Evol. 42 (4), 414-421 (1996)
96215866
Location/Qualifiers
1..3641
/organism="Buchnera aphidicola"
/specific_host="Rhopalosiphum padi"
/db_xref="taxon:9"
1..1545
/gene="trpG"
1..1545
/gene="trpG"
/codon_start-1
/transl_table-1
/product="anthranilate synthase large subunit"
/protein_id="AB05986.1"
/db_xref="PDB:91082003"
/transl_table="MKKSAYPIEIIIOKAPYHPDPYMFNFHLCESRSETLLTADNK
KRSRKMIDMSAMRISSEHNAVKTPLSTINGEILLVYIKTIISKIEIYRRKKNIL
IPFNIEKDDDEKLESYVDFALITTEFERREKSKAMFFGGLFSTDLVSFETL
PKLGNQKSNCFEYLAETLVLDHQQKTLQSSFTNSNEKRIEERSEIYKOL
NOYKLNIPKIKIDINLJNNNFEGSISIKKLOKIGGEIYOVVSRKFLPCBP
LSAYOKLKSNSPWFYQDQDTEFGASPESSLYDKTRIEIYPIAGTRPGRT
KQNDLDDDSRIELEMRTNHELAELMLVLDARNDLARIKCPGSRYSDDLVRVRY
SHMHVSVSVEIGELREGIDALAHAYNSCMNGTITGAPYCANQLIEYERKESYGG
AICYPFDLGNLDCITIRSAIYERKICYNAGAGIYNSIPEDEVNESLKAQAVINA
IKNAH"
1561..2149
/gene="trpG"
1561..2149
/gene="trpG"
/codon_start-2
/transl_table-1
/product="anthranilate synthase small subunit"
/protein_id="AB05987.1"
/db_xref="PDB:91082004"
/db_xref="GI:1082004"
/transl_table="MADILLDNLSFTYNLVEQLKNNNNVLIYRNNIDIKILINSI
NKIKNPILSPGSPKKNAGCMENLINVKGRKIPVIGICLGHQALVEAYGCIIGYAG
EIFHGRASLINDGLEMEFGLPPLPVARYHSLCNKIPKRNINSYFNDMIMSVANN
LIVCGCFQHPSPSILITSCALLLEKIDWASSKY"
BASE COUNT 1490 a 453 c 503 g 1195 t
ORIGIN

Query Match 100.0%; Score 11; DB 1; Length 3641;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGTTCTAGT 11
|||||
DB 121 CTGTTCTAGT 111

```

Search completed: September 17, 1999, 22:01:12  
 Job Time: 16:25 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:33:31 ; Search time 425.19 Seconds

(without alignments)  
6.473 Million cell updates/sec

## ALIGNMENTS

Title: us-08-956-518a-88

Perfect score: 11

Sequence: 1 CTGTTCTACT 11

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database: N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	11	100.0	735	1 N91020
2	11	100.0	5406	1 N93197
3	11	100.0	3763	1 Q06338
4	11	100.0	3316	1 Q05632
5	11	100.0	3688	1 Q05630
6	11	100.0	2863	1 Q05631
7	11	100.0	3824	1 Q1482
8	11	100.0	1920	1 Q13830
9	11	100.0	1592	1 Q14501
10	11	100.0	2232	1 Q06830
11	11	100.0	391	1 Q08935
12	11	100.0	531	1 Q08937
13	11	100.0	5836	1 Q06236
14	11	100.0	5712	1 Q03027
15	11	100.0	24026	1 T18325
16	11	100.0	5914	1 T18310
17	11	100.0	5914	1 T18311
18	11	100.0	5915	1 T18312
19	11	100.0	5914	1 T18313
20	11	100.0	5903	1 T18314
21	11	100.0	153	1 T22290
22	11	100.0	5914	1 T17445
23	11	100.0	5914	1 T17446
24	11	100.0	5912	1 T17447
25	11	100.0	5910	1 T17448
26	11	100.0	5914	1 T17449
27	11	100.0	5913	1 T17450
28	11	100.0	5915	1 T17451
29	11	100.0	5912	1 T17452
30	11	100.0	5914	1 T17453
31	11	100.0	24025	1 T17454
32	11	100.0	5924	1 T17455
33	11	100.0	5914	1 T17456
34	11	100.0	5915	1 T17457
35	11	100.0	5874	1 T17458
36	11	100.0	5915	1 T17459
37	11	100.0	5915	1 T17460
38	11	100.0	5915	1 T17461
39	11	100.0	5913	1 T17462
40	11	100.0	5912	1 T17463
41	11	100.0	5913	1 T17464
42	11	100.0	5914	1 T17465
43	11	100.0	5913	1 T17466
44	11	100.0	5914	1 T17467

C 44	11	100.0	5914	1	T17468	Mutated BRCA1 codi
C 45	11	100.0	597	1	X30620	H. pylori cellular

  

RESULT 1	
ID N91020/c	N91020 standard; DNA; 735 BP.
AC N91020:	
DE 21-JUN-1990 (first entry)	
DE Nicotiana tabacum cDNA for pathogenesis-related (P-R) protein	
KW Pathogenesis-related protein; Nicotiana tabacum cv. Samsun NN;	
KM pNTSNNCPRI/81; pNTSNNCPRI/1.	
OS Nicotiana tabacum cv. Samsun NN.	
PH Key	Location/Qualifiers
FT misc-feature	1..722
FT	/*tag- a
FT	/note="cDNA from pNTSNNCPRI/1"
FT	14..735
FT	/*tag- b
FT	/note="cDNA from pNTSNNCPRI/81"
FT	1..497
FT	/*tag- c
FT	498..735
FT	/*tag- d
PN	EP-307841-A.
PD	22-MAR-1989.
PE	12-SEP-1988; 114877.
PR	15-SEP-1987; US-096579.
PI (GHO-) Gen Hospital Corp.	
PI Pfltzner UM, Pfltzner AP, Goodman HM;	
DR WPI: 89-08735/12.	
DR P-PSDB; P91058.	
PT Pathogenesis-related proteins and regulatory sequences	
PT used for expression in plants to enhance hypersensitive response	
PT to invading pathogen	
PS Figure 3: 33pp; English.	
CC It is cDNA from tobacco mosaic virus (TMV) infected tobacco plants.	
CC It contains the sequence from two independent cDNA clones	
CC pNTSNNCPRI/81 and pNTSNNCPRI/1. Differences between the two clones are	
CC shown in the Features table. Expression of a recombinant DNA sequence	
CC which includes a PR-1 promoter in a plant enhances the hypersensitive	
CC response of the plant to an invading pathogen.	
CC Sequence 735 BP; 227 A; 145 C; 145 G; 218 T;	

  

QY	1 CTGTTCTACT 11	
DB	377 CTGTTCTACT 367	

  

RESULT 2	
ID N93197	N93197 standard; DNA; 5406 BP.
AC N93197:	
DE 28-JUN-1990 (first entry)	
DE Fragment of clone lambda HK65a contg. kallikrein gene.	
KW Kallikrein gene; vasodilator; male infertility; lambda HK65a; ss.	
OS Homo sapiens.	
PH Key	Location/Qualifiers
FT exon	803..848
FT	/*tag- a
FT	2538..2697
FT	/*tag- b
FT	3967..4256
FT	/*tag- c
FT	4375..4334
FT	exon

```

FT      exon                      /*tag= d
FT      4838..5075                /*tag= e
PN      EP-297913-A.
PD      04-JAN-1989.
PF      30-JUN-1988; 306039.
PR      30-JUN-1987; US-068594.
PA      (AMGE-) Amgen Inc.
PI      Lin FK, Lu HS.
DR      WPI; 89-009139/02.
P-PSDB; P93722.
PT      New recombinant kallikrein polypeptide(s) and encoding DNA -
PT      useful as vasodilators and for treating male infertility.
PS      Table V; page 13; 43pp; English.
CC      Lambda HK64a is a subclone produced from clones isolated from a CHA phage
CC      boune human fetal liver genomic library using monkey kallikrein MM680a
CC      clone DNA as a probe.
CC      See also N93193-8.
SQ      Sequence 5406 BP; 974 A; 1864 C; 1309 G; 1259 T;

Query Match
Best Local Similarity 100.0%; Score 11; DB 1; Length 5406;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTGTTCTAGT 11
DB      3532 CTGTTCTAGT 3542

RESULT 3
Q06328/c
AC      Q06328:
ID      006328 standard; DNA; 3763 BP.
DE      06-FEB-1991 (first entry)
DE      Sequence encoding mezoite apical-end-localised protein (MAEP)
DE      Insert 5.3.
KM      Malaria; vaccine; Duffy blood group; ds.
OS      Plasmodium vivax.
PN      WO9011772-A.
PD      18-OCT-1990.
PF      03-APR-1990; U01849.
PR      05-APR-1988; US-334041.
PR      06-APR-1989; US-334270.
PA      (UYNV-) NEW YORK UNIT.
PI      Barnwell JM, Galinski MR, Wertheimer SP;
PI      WPI; 90-334616/44.
P-PSDB; R07503.
PT      Malarial apical end mezoite proteins and peptide(s) - used for
PT      developing cpos. for treating, preventing and diagnosing malarial
PT      infection.
PS      Claim 2; Fig 1a; 66pp; English.
CC      A MAEP compound having a binding affinity for a Duffy blood group
CC      antigen of primate red blood cells, is antigenic for the complete
CC      protein, and may be used in diagnosis, treatment and vaccination
CC      against invasion by P.vivax and P.knowlesi.
SQ      Sequence 3763 BP; 1796 A; 464 C; 611 G; 892 T;

Query Match
Best Local Similarity 100.0%; Score 11; DB 1; Length 3763;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTGTTCTAGT 11
DB      2305 CTGTTCTAGT 2295

RESULT 4
Q05632
AC      Q05632 standard; DNA; 3316 BP.
ID      Q05632:
PI      17-DEC-1990 (first entry)

```

```

DE      Lambda gti0ch101 encodes C-terminal alpha amidating enzyme.
KM      C-terminal alpha amidating enzyme; human thyroid gland;
KW      lambda gti0ch101; ss.
OS      Homo sapiens.
FH      Key
FT      cds
FT      Location/Qualifiers
FT      56..2656
FT      /*tag= a
FT      /product=C-terminal alpha amidating enzyme

QY      1 CTGTTCTAGT 11
DB      2813 CTGTTCTAGT 2823

RESULT 5
Q05630
AC      Q05630 standard; cDNA; 3688 BP.
ID      Q05630:
DE      17-DEC-1990 (first entry)
DE      Lambda gti0ch201 encodes C-terminal alpha amidating enzyme.
DE      C-terminal alpha amidating enzyme; human thyroid gland;
KM      lambda gti0ch201; ss.
OS      Homo sapiens.
FH      Key
FT      cds
FT      Location/Qualifiers
FT      140..3064
FT      /*tag= a
FT      /product=C-terminal alpha amidating enzyme.

QY      1 CTGTTCTAGT 11
DB      2613 CTGTTCTAGT 2623

RESULT 6
Q05631
AC      Q05631 standard; cDNA; 3688 BP.
ID      Q05631:
DE      17-DEC-1990 (first entry)
DE      Lambda gti0ch201 encodes C-terminal alpha amidating enzyme.
DE      C-terminal alpha amidating enzyme; human thyroid gland;
KM      lambda gti0ch201; ss.
OS      Homo sapiens.
FH      Key
FT      cds
FT      Location/Qualifiers
FT      140..3064
FT      /*tag= a
FT      /product=C-terminal alpha amidating enzyme.

QY      1 CTGTTCTAGT 11
DB      2613 CTGTTCTAGT 2623

RESULT 7
Q05632
AC      Q05632 standard; DNA; 3316 BP.
ID      Q05632:
PI      17-DEC-1990 (first entry)
PI      17-JAN-1989; JP-005878.
PA      (SUNR) SUNTORI LTD.
PI      Onsuje K, Kitano K, Tanaka S;
PI      WPI; 90-254034/33.
P-PSDB; R96379.
PT      C-terminal alpha amidating enzymes from human thyroid gland -
PT      converting C-terminal glycy peptide(s) or protein, and new DNA
PT      coding sequences.
PS      Disclosure; pp; English.
CC      cDNA libraries were prepared from human thyroid gland poly(A) RNA
CC      in lambda gti1 and gti0. These were screened using probes derived
CC      from DNA encoding similar enzymes in Xenopus laevis. Three clones
CC      were isolated from the gti0 library, including gti0ch101.
CC      Restriction analysis indicated that gti0ch101 encoded a different
CC      type of enzyme to one of the other two clones. The DNA can be
CC      inserted into vectors for expression in E.coli or (more efficiently)
CC      in animal cells.
CC      The enzymes are useful for prodn. of physiologically active alpha-
CC      amidated peptides from their C-glycyl precursors.
CC      See also Q05630 and Q05631.
SQ      Sequence 3316 BP; 931 A; 676 C; 755 G; 954 T;

Query Match
Best Local Similarity 100.0%; Score 11; DB 1; Length 3316;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTGTTCTAGT 11
DB      2813 CTGTTCTAGT 2823

RESULT 8
Q05633
AC      Q05633 standard; DNA; 3316 BP.
ID      Q05633:
PI      17-DEC-1990 (first entry)
PI      17-JAN-1989; JP-005878.
PA      (SUNR) SUNTORI LTD.
PI      Onsuje K, Kitano K, Tanaka S;
PI      WPI; 90-254034/33.
P-PSDB; R96379.
PT      C-terminal alpha amidating enzymes from human thyroid gland -
PT      converting C-terminal glycy peptide(s) or protein, and new DNA
PT      coding sequences.
PS      Disclosure; pp; English.
CC      cDNA libraries were prepared from human thyroid gland poly(A) RNA
CC      in lambda gti1 and gti0. These were screened using probes derived
CC      from DNA encoding similar enzymes in Xenopus laevis. Three clones
CC      were isolated from the gti0 library, including gti0ch201.
CC      Restriction analysis indicated that gti0ch201 encoded a different

```



CC type of enzyme to the other two clones. The DNA can be inserted  
 CC into vectors for expression in E.coli or (more efficiently) in  
 CC animal cells.  
 CC The enzymes are useful for prodn. of physiologically active alpha-  
 CC amidated peptides from their C-glycyl precursors.  
 CC See also 005631 and 005632.  
 SQ Sequence 3688 BP; 1027 A; 775 C; 862 G; 1024 T;

Query Match 100.0%; Score 11; DB 1; Length 3688;  
 Best Local Similarity 100.0%; Pred. NO. 3.5e+02;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTCTTCTAGT 11  
 DB 3221 CTTCTTCTAGT 3231

631  
 005631 standard; DNA; 2863 BP.

AC 005631;  
 DE 17-DEC-1990 (first entry)  
 DE Lambda gt10ch2 encodes C-terminal alpha amidating enzyme.  
 KM C-terminal alpha amidating enzyme; human thyroid gland;  
 KM Lambda gt10ch2; ss.  
 OS Homo sapiens.  
 FT key Location/Qualifiers  
 FT 2..2332  
 FT /tag= a  
 FT /product=C-terminal alpha amidating enzyme

PN W09008190-A.  
 PD 26-JUL-1990.  
 PF 17-JAN-1990; J00042.  
 PR 17-JAN-1989; JP-005878.  
 PA (SUNR) SUNTORY LTD.  
 PI Ohshye K, Kitanio K, Tanaka S;  
 DR WPI: 90-234034/33.  
 P-PSDB; R96400.

PT C-terminal alpha amidating enzymes from human thyroid gland -  
 PT converting C-terminal glycyl peptide(s) or protein, and new DNA  
 PT coding sequences.

PS Disclosure; pp; English.  
 CC cDNA libraries were prepared from human thyroid gland poly(A) RNA  
 CC in Lambda gtl1 and gtl10. These were screened using probes derived  
 CC from DNA encoding similar enzymes in Xenopus laevis. Three clones  
 CC were isolated from the gtl10 library, including gtl10ch2.  
 CC Restriction analysis indicated that gtl10ch2 encoded a different  
 CC type of enzyme to one of the other two clones. The DNA can be  
 CC inserted into vectors for expression in E.coli or (more efficiently)  
 CC in animal cells.

CC The enzymes are useful for prodn. of physiologically active alpha-  
 CC amidated peptides from their C-glycyl precursors.  
 CC See also 005630 and 005632.  
 CC Sequence 2863 BP; 833 A; 572 C; 656 G; 800 T;

Query Match 100.0%; Score 11; DB 1; Length 2863;  
 Best Local Similarity 100.0%; Pred. NO. 3.5e+02;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTCTTCTAGT 11  
 DB 2489 CTTCTTCTAGT 2499

RESULT 7  
 Q11482  
 ID Q11482 standard; cDNA; 3824 BP.  
 AC Q11482;  
 DE 19-JUN-1991 (first entry)  
 DE Sequence encoding spike protein of infectious bronchitis virus.  
 KM IBV; spike protein; vaccine; ss.

OS Infectious bronchitis virus.  
 FH Key Location/Qualifiers  
 FT cds 151..3657  
 FT /tag= a  
 FT /product= IBV spike protein  
 FT 151..1767

FT misc\_rna  
 FT /tag= b  
 FT /product= IBV spike protein S1 subunit  
 FT 1783..3654  
 FT /tag= c  
 FT /product= IBV spike protein S2 subunit  
 FT 1768..1782  
 FT /tag= d  
 FT /label= joining region

EP-423869-A.  
 PD 20-APR-1991.  
 PF 08-OCT-1990; 202667.  
 PR 20-OCT-1989; US-424793.  
 PA (ALKU) Akzo NV.  
 PI Sondermeijer PJ, Claessens JA;  
 DR WPI: 91-119142/17.  
 P-PSDB; R11611.

PT Nucleic acid sequence - encodes spike protein polypeptide of  
 PT Infectious bronchitis virus  
 PS Disclosure; page 10; 20pp; English.

CC The protein encoded by this sequence is a recombinant spike protein  
 CC of infectious bronchitis virus (IBV) Arkansas strain. This protein  
 CC or an antigenic fragment, e.g. S1 or S2 subunit, can be used in a  
 CC vaccine for immunising poultry against IBV infections.  
 CC See also Q11483-85.

SQ Sequence 3824 BP; 1114 A; 633 C; 733 G; 1344 T;

Query Match 100.0%; Score 11; DB 1; Length 3824;  
 Best Local Similarity 100.0%; Pred. NO. 3.5e+02;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTCTTCTAGT 11  
 DB 166 CTTCTTCTAGT 176

RESULT 8  
 Q13830  
 ID Q13830 standard; DNA; 1920 BP.

AC Q13830;  
 DT 09-DEC-1991 (first entry)  
 DE MSP gene in pUCRS (DSM 5803).  
 KM Major secretion product; expression cassette; ss.  
 OS Lactococcus lactis IM0230 (DSM 5805).

FH Key Location/Qualifiers  
 FT promoter 1..40  
 FT /tag= a  
 FT /note= "part of promoter region"

FT signal\_peptide 411..491  
 FT /tag= b  
 FT mat\_peptide 492..1793  
 FT /tag= c

EP-449770-A.

PD 02-OCT-1991.  
 PF 13-MAR-1991; 810167.  
 PR 22-MAR-1990; GB-006400.  
 PA (CIBA) CIBA GEIGY AG.  
 PI Surti B, Schmitz A;  
 DR WPI: 91-250162/40.  
 P-PSDB; R14150.

PT Hybrid vectors for expression of polypeptide(s) - comprise DNA  
 PT from plasmids obtd. from Lactococcus lactis, esp. major secretion  
 PT prod. gene fragments.

PS Disclosure; page 21; 36pp; English.  
 CC The sequence (SDO ID NO:1) was obtd. from a genomic library of L.  
 CC lactis in Lambda Emu1 3. The promoter region, and opt. the gene,  
 CC can be fused with a gene encoding a protein for secretion of the

CC expressed protein into the supernatant.  
 CC See also Q13831.  
 SQ Sequence 1920 BP; 679 A; 344 C; 357 G; 540 T;

Query Match 100.0%; Score 11; DB 1; Length 1920;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGTTCTAGT 11  
 |||||  
 DB 832 CTGTTCTAGT 842

RESULT 9  
 Q14501  
 ID Q14501 standard; DNA; 1592 BP.  
 AC Q14501;  
 PM 28-JAN-1992 (first entry)

USP45 gene.  
 Chymosin; alpha-amylase; PN21011; enzyme; ss.  
 OS Lactococcus lactis spp. lactis MG1363.  
 FH Key Location/Qualifiers

FT cds 101..1483  
 FT /\*tag- a  
 FT /label- dsp45  
 FT signal\_peptide 101..181  
 FT /\*tag- b  
 FT mat\_peptide 182..1483  
 FT /\*tag- c  
 FT rbs 75..79  
 FT /\*tag- d  
 FT rbs 90..93  
 FT /\*tag- e  
 FT -10\_signal 45..50  
 FT /\*tag- f  
 FT -35\_signal 25..30  
 FT /\*tag- g  
 FT misc\_signal 57  
 FT /\*tag- h  
 FT /note- "start of transcription"  
 FT repeat\_unit 1514..1553  
 FT /\*tag- i  
 FT /rpl\_type- inverted  
 FT /note- "rho-independent terminator"

PN N19000753-A.  
 16-OCT-1991.  
 30-MAR-1991; 200721.  
 (NEZU-) NEDERLANDS INST ZUI.

PA WPI; 91-330695/45.  
 DR P-PSDB; R14530.

PT Lactococcus lactis DNA fragments - conty. extracellular protein  
 PT signal peptide sequence  
 PS Disclosure; Fig 4; 20pp; Dutch.

CC The sequence is the KpnI/ClaI fragment from PN21011. Lactococci  
 CC transformed with the gene may be used to produce extracellular  
 CC heterologous enzymes of use in the food industry, e.g. chymosin or  
 CC alpha-amylase.

CC See also Q14502.  
 SQ Sequence 1592 BP; 579 A; 291 C; 308 G; 414 T;

Query Match 100.0%; Score 11; DB 1; Length 1592;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGTTCTAGT 11  
 |||||  
 DB 522 CTGTTCTAGT 532

RESULT 10

Q060820/C  
 ID Q060820 standard; DNA; 391 BP.  
 AC Q060820;  
 DR 16-MAR-1994 (first entry)

DE Human brain Expressed Sequence Tag EST00924.  
 KW Gene transcription product; genetic markers; tagging; in vivo;  
 KW transcription; mapping; locations; chromosomes; chromosomal; ss.  
 OS Homo sapiens.

PN W09316178-A.  
 19-AUG-1993.  
 PF 12-FEB-1993; U01294.  
 PR 12-FEB-1992; US-837195.

PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
 PI Adams MD, Moreno RF, Venter CJ;  
 DR WPI; 93-272882/34.

PT Enriched oligonucleotides and corresp. sequences - used as  
 PT markers for human genes transcribed in-vivo, facilitate tagging  
 PT of most human genes

PS Example 4; Page 395-396; 500pp; English.  
 CC The Expressed Sequence Tag was isolated from a human brain CDNA  
 CC library as part of a large set of ESTs which can be used as markers  
 CC for human genes transcribed in vivo. They can be used to facilitate  
 CC tagging of most human genes, for mapping locations of expressed genes  
 CC on chromosomes, for individual or forensic identification, for mapping  
 CC locations of disease-associated genes, for identification of tissue  
 CC type, and for prep. of antisense sequences, probes and constructs.

CC EST00924 has a "poor" coding probability as evaluated using the  
 CC coding-region prediction program CRM. See also Q59041-Q61440.  
 SQ Sequence 391 BP; 97 A; 106 C; 115 G; 71 T;

Query Match 100.0%; Score 11; DB 1; Length 391;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGTTCTAGT 11  
 |||||  
 DB 76 CTGTTCTAGT 66

RESULT 11

Q08935/C  
 ID Q08935 standard; CDNA; 2232 BP.  
 AC Q08935;  
 DR 04-DEC-1995 (first entry)

DE Human death associated protein DAP-1.  
 KW Death associated protein; DAP; cytokine; cell death; ss.  
 OS Homo sapiens.

FH Key Location/Qualifiers  
 FT cds 160..468  
 FT /\*tag- a  
 FT polya\_signal 2203..2208  
 FT /\*tag- b  
 FT misc\_feature 160..468  
 FT /\*tag- c  
 FT /note- "Claimed"

FT misc\_feature 287..818  
 FT /\*tag- d  
 FT /note- "Claimed"

PN W09510630-A.  
 20-APR-1995.  
 PD 12-OCT-1994; U11598.

PR 12-OCT-1993; IL-107250.  
 PA (RYCU/) RYCUS A.  
 PA (YEDA ) YEDA RES & DEV CO LTD.

PI Kimchi A;  
 DR WPI; 95-178528/23.

DR P-PSDB; R74204.  
 PT DNA whose expression mediates cytokine-induced programmed cell  
 PT death - used to treat diseases or disorders associated with  
 PT uncontrolled, pathological cell growth or cytokine-induced  
 PT programmed cell death.  
 PS Claim 2; Fig 6; 61pp; English.

CC DAP genes seem to play an imp. role in programmed cell death and the  
 CC inhibition of their expression protects the cell from cytokine-  
 CC promoted cell death. An HL-60 cDNA library constructed in lambda  
 CC 9110 vector was screened with the cDNA insert of PTK01-230. Two  
 CC independent clones, lambda1 and lambda2, almost completely  
 CC overlapping and carrying cDNA inserts of about 2.3 kb were analysed.  
 CC Lambda1 cDNA clone encompasses the 5'-UTR, short coding region(s)  
 CC and a relatively long 3' UTR that constitutes more than 60% of the  
 CC cDNA clone. The nt. sequence of the cDNA carried by lambda1 and  
 CC predicted AA pattern are present in Q89835/R74204. The ORF is  
 CC preceded by an extremely GC-rich 5' UTR and potentially codes for  
 CC a protein consisting of 102 AAs with calculated MW of 11.2 kDa.  
 CC It has an isoelectric pt. of 10 and is rich in prolines.  
 CC Sequence 2232 BP; 542 A; 647 C; 569 G; 474 T;  
 SQ

Query Match 100.0%; Score 11; DB 1; Length 2232;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CTGTTCTAGT 11  
 |||||  
 Db 636 CTGTTCTAGT 626

RESULT 12  
 ID Q89837/c 089837 standard; cDNA; 531 BP.  
 AC Q89837; standard; cDNA; 531 BP.  
 DT 04-DEC-1995 (first entry)  
 DE Human death associated protein DAP-1.  
 KW Death associated protein; DAP; cytokine; cell death; ss.  
 OS Homo sapiens.  
 FH Key location/Qualifiers  
 FT cds 1..182  
 FT /\*tag- a  
 FN W09510630-A.  
 PD 20-APR-1995.  
 PF 12-OCT-1994; U11598.  
 PR 12-OCT-1993; IL-107250.  
 PA (RVCU/) RVCUS A.  
 PA (YEDA ) YEDA RES & DEV CO LTD.  
 PI Klmch1 A;  
 DR WPI; 95-178528/23.  
 PT DNA whose expression mediates cytokine-induced programmed cell  
 PT death - used to treat diseases or disorders associated with  
 PT uncontrolled, pathological cell growth or cytokine-induced  
 PT programmed cell death.  
 PT Claim 2; Fig 6; 61p; English.  
 CC DAP genes seem to play an imp. role in programmed cell death and the  
 CC inhibition of their expression protects the cell from cytokine-  
 CC promoted cell death. An HL-60 cDNA library constructed in lambda  
 CC 9110 vector was screened with the cDNA insert of PTK01-230. Two  
 CC independent clones, lambda1 and lambda2, almost completely  
 CC overlapping and carrying cDNA inserts of about 2.3 kb were analysed.  
 CC Lambda1 cDNA clone encompasses the 5'-UTR, short coding region(s)  
 CC and a relatively long 3' UTR that constitutes more than 60% of the  
 CC cDNA clone. The nt. sequence of the cDNA carried by lambda1 and  
 CC predicted AA pattern are present in Q89835/R74204. The ORF is  
 CC preceded by an extremely GC-rich 5' UTR and potentially codes for  
 CC a protein consisting of 102 AAs with calculated MW of 11.2 kDa.  
 CC It has an isoelectric pt. of 10 and is rich in prolines. Q89837  
 CC comprises a part of the ORF of Q89835 plus a part of the 3' UTR.  
 CC Sequence 531 BP; 134 A; 153 C; 119 G; 125 T;  
 SQ

Query Match 100.0%; Score 11; DB 1; Length 531;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTAGT 11  
 |||||  
 Db 349 CTGTTCTAGT 339

RESULT 13  
 ID Q96296/c 096296 standard; DNA; 5836 BP.  
 AC Q96296; standard; DNA; 5836 BP.  
 DT 30-DEC-1995 (first entry)  
 DE Human prostate-specific transcriptional regulatory element.  
 KW Transcriptional regulatory element; prostate-specific antigen;  
 KW prostate-specific enhancer; promoter; ss.  
 OS Homo sapiens.  
 FH Key location/Qualifiers  
 FT enhancer 503..2974  
 FT /\*tag- a  
 FT /label- upstream PSA enhancer  
 FT /note- "claimed, nts -5322 - -2851"  
 FT misc\_signal 1..5824  
 FT /\*tag- b  
 FT /label- transcriptional regulatory element  
 FT /note- "nts -5824 - -1"  
 FT misc\_rna 5825..5831  
 FT /\*tag- c  
 FT /label- transcriptional start site  
 FT /note- "nts +1 - +7"  
 FT promoter 5284..5831  
 FT /\*tag- d  
 FT /label- naturally-occurring PSA promoter  
 FT /note- "nts -541 - +7"  
 FT promoter 5506..5831  
 FT /\*tag- e  
 FT /label- naturally-occurring PSA promoter  
 FT /note- "nts -320 - +7"  
 FT misc\_signal 524..3025  
 FT /\*tag- f  
 FT /label- transcriptional regulatory element  
 FT /note- "claimed, nts -5300 - -2800"  
 FT enhancer 503..5824  
 FT /\*tag- g  
 FT /label- transcriptional regulatory element  
 FT /note- "claimed, nts 0 - -5322"  
 FT

PN W09519434-A.  
 PD 20-JUL-1995.  
 PF 12-JAN-1994; U00845.  
 PR 13-JAN-1994; US-182247.  
 PA (CALY-) CALYDON INC.  
 PI Henderson DR.  
 DR WPI; 95-263864/34.  
 PT Transcriptional regulator specific for cells expressing prostate specific  
 PT antigen - used to express toxins, immunostimulants or anti-sense cpds.  
 PT for treatment and prevention of prostatic cancer or hypertrophy.  
 PS Claim 2; Fig 1; 71p; English.  
 CC Q96296 shows the sequence of the 5' flanking region of the human  
 CC prostate specific antigen to -5824 bp. The fragment runs from a  
 CC HindIII site at -5824 bp to the HindIII site at +7 bp. The  
 CC numbering system is +1 at the transcriptional start site of PSA mRNA  
 CC (Lundvall, A., 1989, Biochem. Biophys. Res. Comm. 161: 1151-59). The  
 CC coding region of PSA starts at +42. An approx. 2.5 kbp fragment from  
 CC about -2850 to about -5350 serves as a prostate specific enhancer  
 CC (PSE) in conjunction with a promoter to function to initiate  
 CC transcription in prostate cells. This can be isolated from the  
 CC human genome as a XbaI-ApaI 2.5 kb fragment and this enhancer is  
 CC termed the 'upstream PSA enhancer'. PSA - prostate-specific  
 CC antigen. There is a naturally-occurring PSA promoter spanning  
 CC the segment from about -541 to +7, particularly -320 to +7 of  
 CC the human PSA gene. The promoter of PSA has been reported  
 CC (Riegman et al., 1989, Biochem. Biophys. Res. Comm. 159: 95).  
 CC The promoter from -320 to +7 contains a TATA-box, a GC-box and a  
 CC hormonal response element at -170 to -156.  
 CC Sequence 5836 BP; 1408 A; 1296 C; 1568 G; 1564 T;  
 SQ

Query Match 100.0%; Score 11; DB 1; Length 5836;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;

OS Homo sapiens.

Key	Location/Qualifiers
Intron	1..55
exon	/*tag= a
intron	56..155
exon	/*tag= b
intron	156..1512
exon	/*tag= c
intron	1513..1611
exon	/*tag= d
intron	1612..2206
exon	/*tag= e
intron	/note= "n at 1925-1937 represent an indefinite interval within the intron"
exon	2207..2260
intron	/*tag= f
exon	2261..2677
intron	/*tag= g
exon	/note= "n at 2569-2581 represent an indefinite interval within the intron"
intron	2678..2788
exon	/*tag= h
intron	2789..3328
exon	/*tag= i
intron	/note= "n at 3063-3075 represent an indefinite interval within the intron"
exon	3329..3406
intron	/*tag= j
exon	3407..3813
intron	/*tag= k
exon	/note= "n at 3598-3610 represent an indefinite interval within the intron"
intron	3814..3902
exon	/*tag= l
intron	3903..4224
exon	/*tag= m
intron	/note= "n at 4076-4088 represent an indefinite interval within the intron"
exon	4225..4364
intron	/*tag= n
exon	4365..6571
intron	/*tag= o
exon	/note= "n at 4602-4614 represent an indefinite interval within the intron"
intron	6572..6677
exon	/*tag= p
intron	6678..9163
exon	/*tag= q
intron	9164..9207
exon	/*tag= r
intron	9208..10530
exon	/*tag= s
intron	10531..10607
exon	/*tag= t
intron	10608..11597
exon	/*tag= u
intron	/note= "n at 11383-11396 represent an indefinite interval within the intron"
exon	11598..15023
intron	/*tag= v
exon	15024..15424
intron	/*tag= w
exon	15425..15511
intron	/*tag= x
exon	15512..15952
intron	/*tag= y
exon	/note= "n at 15647-15659 represent an indefinite interval within the intron"
intron	15953..16126
exon	/*tag= z
intron	16127..16565
exon	/*tag= aa
intron	/note= "n at 16370-16382 represent an indefinite interval within the intron"

```

FT      exon      16566..16692
FT      /tag- ab
FT      Intron    16693..17535
FT      /tag- ac
FT      /note- "n at 17290-17302 represent an indefinite
FT      interval within the intron"
FT      exon      17536..17726
FT      /tag- ad
FT      Intron    17727..18416
FT      /tag- ae
FT      /note- "n at 18299-18312 represent an indefinite
FT      interval within the intron"
FT      exon      18417..18787
FT      /tag- af
FT      Intron    18788..19298
FT      /tag- ag
FT      /note- "n at 18952-18964 represent an indefinite
FT      interval within the intron"
FT      exon      19299..19386
FT      /tag- ah
FT      Intron    19387..20190
FT      /tag- ai
FT      /note- "n at 19887-19899 represent an indefinite
FT      interval within the intron"
FT      exon      20191..20267
FT      /tag- aj
FT      Intron    20268..21094
FT      /tag- ak
FT      /note- "n at 20767-20779 represent an indefinite
FT      interval within the intron"
FT      exon      21095..21135
FT      /tag- al
FT      Intron    21136..21583
FT      /tag- am
FT      /note- "n at 21341-21353 represent an indefinite
FT      interval within the intron"
FT      exon      21584..21667
FT      /tag- an
FT      Intron    21668..22233
FT      /tag- ao
FT      /note- "n at 21921-21933 represent an indefinite
FT      interval within the intron"
FT      exon      22234..22288
FT      /tag- ap
FT      Intron    22289..22832
FT      /tag- aq
FT      /note- "n at 22567-22579 represent an indefinite
FT      interval within the intron"
FT      exon      22833..22906
FT      /tag- ar
FT      Intron    22907..23287
FT      /tag- as
FT      /note- "n at 23050-23062 represent an indefinite
FT      interval within the intron"
FT      exon      23288..23348
FT      /tag- at
FT      Intron    23349..23698
FT      /tag- au
FT      /note- "n at 23580-23592 represent an indefinite
FT      interval within the intron"
FT      exon      23699..24026
FT      /tag- av
FT      Intron    24027..24725
FT      /tag- aw
FT      /note- "polymorphic site"
FT      Intron    24726..25363
FT      /tag- ax
FT      /note- "polymorphic site"
FT      Intron    25364..25991
FT      /tag- ay
FT      /note- "polymorphic site"
FT      Intron    25992..26439
FT      /tag- az

```

```

FT      misc_feature 6823 /note- "polymorphic site"
FT      /tag- ba
FT      /note- "polymorphic site"
FT      misc_feature 9106 /note- "polymorphic site"
FT      /tag- bb
FT      /note- "polymorphic site"
FT      misc_feature 9207 /tag- bc
FT      /note- "polymorphic site"
FT      misc_feature 9376 /tag- bd
FT      /note- "polymorphic site"
FT      misc_feature 11908 /tag- be
FT      /note- "polymorphic site"
FT      misc_feature 11994 /tag- bf
FT      /note- "polymorphic site"
FT      misc_feature 12952 /tag- bg
FT      /note- "polymorphic site"
FT      misc_feature 13004 /tag- bh
FT      /note- "polymorphic site"
FT      misc_feature 13009 /tag- bi
FT      /note- "polymorphic site"
FT      misc_feature 13048 /tag- bj
FT      /note- "polymorphic site"
FT      misc_feature 13238 /tag- bk
FT      /note- "polymorphic site"
FT      misc_feature 13448 /tag- bl
FT      /note- "polymorphic site"
FT      misc_feature 13539 /tag- bm
FT      /note- "polymorphic site"
FT      misc_feature 13951 /tag- bn
FT      /note- "polymorphic site"
FT      misc_feature 14041 /tag- bo
FT      /note- "polymorphic site"
FT      misc_feature 14046 /tag- bz

```

```

Query Match      100.0%; Score 11; DB 1; Length 24026;
Best Local Similarity 100.0%; Pred.No. 3.6e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 CTTCTTCTAGT 11
DB      13139 CTTCTTCTAGT 13129

```

Search completed: September 18, 1999, 00:33:33  
Job time: 18973 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:27:57 : Search time 2825.05 Seconds  
(without alignments)  
7.681 Million cell updates/sec

Title: US-08-956-518a-88  
Perfect score: 11  
Sequence: 1 CTGTTCTACT 11

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database:

EST:  
1: em\_est1.\*  
2: em\_est2.\*  
3: em\_est3.\*  
4: em\_est4.\*  
5: em\_est5.\*  
6: em\_est6.\*  
7: em\_est7.\*  
8: em\_est8.\*  
9: em\_est9.\*  
10: em\_est10.\*  
11: em\_est11.\*  
12: em\_est12.\*  
13: em\_est13.\*  
14: em\_est14.\*  
15: em\_est15.\*  
16: em\_est16.\*  
17: em\_est17.\*  
18: em\_est18.\*  
19: em\_est19.\*  
20: em\_est20.\*  
21: em\_est21.\*  
22: em\_est22.\*  
23: em\_est23.\*  
24: em\_est24.\*  
25: em\_est25.\*  
26: em\_est26.\*  
27: em\_est27.\*  
28: em\_est28.\*  
29: em\_est29.\*  
30: em\_est30.\*  
31: em\_est31.\*  
32: em\_est32.\*  
33: em\_est33.\*  
34: em\_est34.\*  
35: em\_est35.\*  
36: em\_est36.\*  
37: em\_est37.\*  
38: em\_est38.\*  
39: em\_est39.\*  
40: em\_est40.\*  
41: em\_est41.\*  
42: em\_est42.\*  
43: em\_est43.\*  
44: em\_est44.\*  
45: em\_est45.\*  
46: em\_est46.\*  
47: em\_est47.\*  
48: em\_est48.\*  
49: em\_est49.\*  
50: em\_est50.\*  
51: em\_est51.\*  
52: em\_est52.\*  
53: em\_est53.\*

54: em\_est22.\*  
55: em\_est23.\*  
56: em\_est24.\*  
57: em\_est25.\*  
58: em\_est26.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	11	100.0	195	20	L26727	L26727 MUSF200A.1a
2	11	100.0	391	20	M78776	M78776 EST00924.H1
3	11	100.0	369	20	T08787	T08787 EST06679.In
4	11	100.0	250	20	T10579	T10579 hbc233.Huma
5	11	100.0	183	20	T17272	T17272 N18615.Norm
6	11	100.0	234	20	T18606	T18606 hbc2266.Hum
7	11	100.0	265	20	T28224	T28224 EST32443.Hu
8	11	100.0	428	20	T36400	T36400 EST101330.S
9	11	100.0	346	20	T36417	T36417 EST101348.S
10	11	100.0	328	20	T48192	T48192 yb35h05.r1
11	11	100.0	294	20	T57378	T57378 yb51a04.r1
12	11	100.0	418	20	T57423	T57423 yb51e06.r1
13	11	100.0	513	20	T57520	T57520 yb58b01.s1
14	11	100.0	383	20	Z35023	Z35023 ATTS3662.AC
15	11	100.0	295	20	Z38728	Z38728 HSC08082.n
16	11	100.0	267	20	Z38766	Z38766 HSC00D122.n
17	11	100.0	336	21	F01159	F01159 HSB86F102.S
18	11	100.0	338	21	F04309	F04309 HSC3PD092.n
19	11	100.0	327	21	F04405	F04405 HSC2S8082.n
20	11	100.0	304	21	F04424	F04424 HSC2SFC032.n
21	11	100.0	280	21	F08472	F08472 HSC2YB061.n
22	11	100.0	358	21	F09540	F09540 HSC33A032.n
23	11	100.0	328	21	R04060	R04060 PK18b05.r1
24	11	100.0	392	21	R04483	R04483 PK24c04.r1
25	11	100.0	361	21	R06633	R06633 yf10b02.r1
26	11	100.0	342	21	R10668	R10668 yf31g10.s1
27	11	100.0	385	21	R12759	R12759 yf58h04.r1
28	11	100.0	436	21	R15422	R15422 y448g08.r1
29	11	100.0	518	21	T65394	T65394 yC73c03.s1
30	11	100.0	334	21	T68242	T68242 yC40e09.r1
31	11	100.0	532	21	T77495	T77495 yC94g01.r1
32	11	100.0	283	21	T77506	T77506 yC84g12.r1
33	11	100.0	259	21	T87738	T87738 y608d10.r1
34	11	100.0	343	22	H11107	H11107 ym14d05.s1
35	11	100.0	431	22	R36022	R36022 y669b06.r1
36	11	100.0	328	22	R45658	R45658 y444e10.s1
37	11	100.0	398	22	R55801	R55801 y692b01.s1
38	11	100.0	165	22	R57987	R57987 F7919.Fetal
39	11	100.0	391	22	R66309	R66309 y130g12.r1
40	11	100.0	431	22	R66315	R66315 y130h12.r1
41	11	100.0	328	22	R70708	R70708 y141h11.s1
42	11	100.0	326	22	R74865	R74865 MDB0185R.Mo
43	11	100.0	234	23	D56671	D56671 HUM219C04B
44	11	100.0	248	23	D58198	D58198 HUM355A08B
45	11	100.0	439	54	HSN008808	AL043958 Homo Sapi

ALIGNMENTS

RESULT 1  
L26727 L26727 195 bp mRNA  
LOCUS MUSF200A lambda unizap male testis Mus musculus cDNA clone F200,  
DEFINITION mRNA sequence.  
ACCESSION L26727  
NID L26727  
VERSION L26727.1 GI:437510

```

KEYWORDS      EST.
SOURCE         house mouse.
ORGANISM       Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS        Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE          1 (bases 1 to 195)
JOURNAL        Kerr, S.M., Vambrie, S., McKay, S.J. and Cooke, H.J.
MEDLINE        Analysis of cDNA sequences from mouse testis
COMMENT        Mamm. Genome 5, 557-565 (1994)
               On Apr 14, 1993 this sequence version replaced.

FEATURES
  source
    Contact: Kerr, S.M., Vambrie, S., McKay, S.J. and Cooke, H.J.
    Location/Qualifiers
      1..195
        /organism="Mus musculus"
        /strain="Swiss"
        /db_xref="taxon:10090"
        /clone="F200"
        /note="Vector: lambda unizap; lambda unizap library from
        mail mouse testis"
BASE COUNT    44 a 34 c 44 g 71 t 2 others
ORIGIN
Query Match   100.0%; Score 11; DB 20; Length 195;
Best Local Similarity 100.0%; Pred. No. 3e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTACT 11
Db 37 CTGTTCTACT 47

RESULT 2
LOCUS M78776 391 bp mRNA EST 11-JAN-1995
DEFINITION EST00924 Hippocampus, Strata gene (cat. #936205) Homo sapiens cDNA
clone HHCME73, mRNA sequence.
ACCESSION M78776
NID 9273091
VERSION M78776.1 GI:273091
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE      Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS        1 (bases 1 to 391)
TITLE          Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M.,
               Uterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C.
               Sequence identification of 2,375 human brain genes
JOURNAL        Nature 355, 632-634 (1992)
MEDLINE        92168112
REFERENCE      2 (bases 1 to 391)
AUTHORS        Durbin, A.S.
TITLE          EST mapping data
JOURNAL        Unpublished (1996)
COMMENT        Contact: Kerlavage, AR
               Bioinformatics
               The Institute for Genomic Research
               9712 Medical Center Drive, Rockville, MD 20850 USA
               Tel: 3018699056
               Fax: 3018699423
               Email: arkerlav@tigr.org
               Seq primer: M13 Forward.
               Location/Qualifiers
                 1..391
                   /organism="Homo sapiens"
                   /db_xref="ATCC (inhost):78208"
                   /db_xref="GDB:D051760E"
                   /db_xref="taxon:9606"
FEATURES
  source
    Location/Qualifiers
      1..391
        /organism="Homo sapiens"
        /db_xref="ATCC (inhost):78208"
        /db_xref="GDB:D051760E"
        /db_xref="taxon:9606"

```

```

/map="22"
/clone="HHCME73"
/clone_lib="Hippocampus, Strata gene (cat. #936205)"
/note="Vector: lambdaZAP-II; Female, 2 years; oligo-dT +
random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb
average insert size."
BASE COUNT    97 a 106 c 115 g 71 t 2 others
ORIGIN
Query Match   100.0%; Score 11; DB 20; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.7e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTACT 11
Db 266 CTGTTCTACT 276

RESULT 4
LOCUS T10579 250 bp mRNA EST 29-NOV-1993
DEFINITION hbc233 Human pancreatic islet Homo sapiens cDNA clone hbc233 5'end,
mRNA sequence.
ACCESSION T10579
NID 9390733

FEATURES
  source
    Contact: Adams, MD
    The Institute for Genomic Research
    932 Clippert Road, Galthersburg, MD 20878
    Tel: 3018699056
    Fax: 3018699423
    Email: mdadams@tigr.org
    Seq primer: M13 Reverse.
    Location/Qualifiers
      1..369
        /organism="Homo sapiens"
        /db_xref="ATCC (inhost):85285"
        /db_xref="taxon:9606"
        /clone="H18B71"
        /clone_lib="Infant Brain, Bento Soares"
BASE COUNT    60 a 119 c 97 g 92 t 1 others
ORIGIN
Query Match   100.0%; Score 11; DB 20; Length 369;
Best Local Similarity 100.0%; Pred. No. 2.7e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTACT 11
Db 266 CTGTTCTACT 276

RESULT 3
LOCUS T08787 369 bp mRNA EST 03-AUG-1993
DEFINITION EST06679 Infant Brain, Bento Soares Homo sapiens cDNA clone H18B71
5'end, mRNA sequence.
ACCESSION T08787
NID 9389815
VERSION T08787.1 GI:389815
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE      Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS        1 (bases 1 to 369)
TITLE          Adams, M.D., Soares, M.B., Kerlavage, A.R., Fields, C. and Venter, J.C.
               Rapid cDNA sequencing (expressed sequence tags) from a
               directionally cloned human infant brain cDNA library
JOURNAL        Nature Genet. 4, 373-380 (1993)
MEDLINE        94004965
COMMENT        Contact: Adams, MD
               The Institute for Genomic Research
               932 Clippert Road, Galthersburg, MD 20878
               Tel: 3018699056
               Fax: 3018699423
               Email: mdadams@tigr.org
               Seq primer: M13 Reverse.
               Location/Qualifiers
                 1..369
                   /organism="Homo sapiens"
                   /db_xref="ATCC (inhost):85285"
                   /db_xref="taxon:9606"
                   /clone="H18B71"
                   /clone_lib="Infant Brain, Bento Soares"

```

VERSION T10579.1 GI:390733  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 250)  
 AUTHORS Takeda, J., Yano, H., Eng, S., Zeng, Y. and Bell, G. I.  
 TITLE A molecular inventory of human pancreatic islets: sequence analysis  
 of 1000 cDNA clones  
 JOURNAL Hum. Mol. Genet. 2, 1793-1798 (1993)  
 MEDLINE 94108427  
 COMMENT Contact: Bell GI or Takeda J  
 HHMI  
 Univ. of Chicago  
 5841 S. Maryland Ave., MC1028, Chicago IL 60637  
 Tel: 3127029116  
 Fax: 3127020271  
 Email: g-bell@uchicago.edu  
 Seq primer: SK primer.  
 Location/Qualifiers  
 1..250  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="hbc233"  
 /note="Vector: Lambda ZAPII; Site\_1: Eco RI; Site\_2: Xho  
 I; mRNA was prepared from normal adult human islets. cDNA  
 was directionally synthesized from the Xho I in the vector  
 to the EcoRI site. cDNA was size fractionated to remove  
 sequences <1000 bp in size."  
 BASE COUNT 60 a 47 c 63 g 80 t  
 ORIGIN  
 Query Match 100.0%; Score 11; DB 20; Length 250;  
 Best Local Similarity 100.0%; Pred. No. 2.9e+03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTGTTCTAGT 11  
 |||||  
 Db 82 CTGTTCTAGT 92  
 RESULT 5  
 --7272/c  
 T17272 183 bp mRNA EST 25-JUL-1996  
 NIB615 Normalized Infant brain, Bento Soares Homo sapiens CDNA  
 3' end, mRNA sequence.  
 ACCESSION T17272  
 NID 9519434  
 VERSION T17272.1 GI:519434  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 183)  
 AUTHORS Berry, R., Stevens, T. J., Walter, N. A. R., Wilcox, A. S., Rubano, T.,  
 Hopkins, J. A., Weber, J., Gould, R., Soares, M. B. and Sikele, J. M.  
 TITLE Gene-based Sequence Tagged Sites (STSS) as the basis for a human  
 gene map  
 JOURNAL Nature Genet. 10, 415-423 (1995)  
 MEDLINE 95400322  
 COMMENT Contact: Sikele JM  
 Department of Pharmacology  
 University of Colorado Health Sciences Center  
 Box C236, 4200 E. 9th Ave, Denver CO 80262-0236  
 Tel: 3032708637  
 Fax: 3032707097  
 Email: nikki@tally.uchsc.edu

FEATURES Seq primer: -21M13 Universal.  
 Location/Qualifiers  
 1..183  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="hbc2266"  
 /note="Vector: EcoRI; Site\_1: HindIII; Site\_2:  
 NotI; The normalized infant brain library, constructed by  
 Bento Soares, Columbia University, was oligo-(dT) primed  
 and directionally cloned into an M13-derived plasmid using  
 total brain mRNA from a 72-day old human female afflicted  
 with spinal muscular atrophy. The library was normalized  
 as described elsewhere."  
 BASE COUNT 73 a 29 c 26 g 50 t 5 others  
 ORIGIN  
 Query Match 100.0%; Score 11; DB 20; Length 183;  
 Best Local Similarity 100.0%; Pred. No. 3e+03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTGTTCTAGT 11  
 |||||  
 Db 44 CTGTTCTAGT 34  
 RESULT 6  
 T18606 234 bp mRNA EST 28-APR-1994  
 LOCUS hbc2266 Human pancreatic islet Homo sapiens CDNA clone hbc2266 5',  
 mRNA sequence.  
 ACCESSION T18606  
 NID 9475414  
 VERSION T18606.1 GI:475414  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 234)  
 AUTHORS Bell, G. I. and Takeda, J.  
 TITLE Human pancreatic islet CDNAS  
 JOURNAL Unpublished (1995)  
 COMMENT On Sep 21, 1992 this sequence version replaced gi:276118.  
 Contact: Bell GI or Takeda J  
 HHMI  
 Univ. of Chicago  
 5841 S. Maryland Ave., MC1028, Chicago IL 60637  
 Tel: 3127029116  
 Fax: 3127020271  
 Email: g-bell@uchicago.edu  
 Seq primer: SK primer.  
 Location/Qualifiers  
 1..234  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="hbc2266"  
 /note="Vector: Lambda ZAPII; Site\_1: Eco RI; Site\_2: Xho  
 I; mRNA was prepared from normal adult human islets. cDNA  
 was directionally synthesized from the Xho I in the vector  
 to the EcoRI site. cDNA was size fractionated to remove  
 sequences <1000 bp in size."  
 BASE COUNT 79 a 46 c 40 g 69 t  
 ORIGIN  
 Query Match 100.0%; Score 11; DB 20; Length 234;  
 Best Local Similarity 100.0%; Pred. No. 2.9e+03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 1 CTGTTCTAGT 11  
 |||||  
 DB 82 CTGTTCTAGT 92

RESULT 7  
 T28224 265 bp mRNA EST 06-SEP-1995  
 LOCUS T28224  
 DEFINITION EST32443 Human Embryo Homo sapiens CDNA 5' end similar to  
 peptidylglycine alpha-amidating monooxygenase (HT:1157), mRNA  
 sequence.  
 ACCESSION T28224  
 NID 9610322  
 VERSION T28224.1 GI:610322  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 265)  
 Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fudner,R.A.,  
 Bul,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
 White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,  
 Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
 Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Georghagen,N.S.M.,  
 Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkley,F.S.,  
 Kelley,J.M., Klimke,K.M., Kelley,J.C., Liu,L.-I., Matmaros,S.M.,  
 Merriek,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T.,  
 Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,  
 Sauder,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R.,  
 Weidman,T.F., Li,Y., Bednarek,D.P., Ceol,L., Cepeda,M.A.,  
 Coleman,T.A., Collins,E.-D., Dimke,D., Feng,P., Ferrite,A.,  
 Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,  
 Giber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Li,H., Li,H.,  
 Meisner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wang,J., Xu,C.,  
 Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,  
 Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.  
 Initial Assessment of Human Gene Diversity and Expression Patterns  
 Based Upon 83 Million Basepairs of CDNA Sequence  
 Nature 377, 3-174 (1995)  
 96026280

TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT

Contact: Venter, J.C.  
 The Institute for Genomic Research  
 932 Clopper Rd, Gathersburg, MD 20878  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: tdbinfo@tdb.tigr.org  
 For clone availability, additional sequence and expression  
 information related to this EST, please contact the TIGR Database  
 (tdbinfo@tdb.tigr.org)  
 Seq primer: M13 Reverse.

FEATURES  
 Location/Qualifiers  
 source 1..265  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):101601"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Human Embryo"  
 /tissue\_type="Embryo"

BASE COUNT 68 a 52 c 43 g 101 t 1 others

ORIGIN

Query Match 100.0%; Score 11; DB 20; Length 265;  
 Best Local Similarity 100.0%; Pred. No. 2.9e+03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTAGT 11  
 |||||  
 DB 52 CTGTTCTAGT 62

RESULT 8

T36400  
 LOCUS T36400 428 bp mRNA EST 11-JAN-1995  
 DEFINITION EST101330 S. cerevisiae strain X2180-1A Saccharomyces cerevisiae  
 CDNA 3' end, mRNA sequence.  
 ACCESSION T36400  
 NID 9620217  
 VERSION T36400.1 GI:620217  
 KEYWORDS EST.  
 SOURCE Baker's yeast.  
 ORGANISM Saccharomyces cerevisiae  
 Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 Saccharomycetaceae; Saccharomyces.  
 1 (bases 1 to 428)  
 Weinstock,K.  
 Saccharomyces cerevisiae CDNAs  
 Unpublished (1995)  
 Other ESTs: TC2  
 Contact: Weinstock,K. and Venter,J.C.  
 The Institute for Genomic Research  
 932 Clopper Rd, Gathersburg, MD 20878  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: tdbinfo@tdb.tigr.org  
 For clone availability please contact the TIGR Database  
 (tdbinfo@tdb.tigr.org)  
 Seq primer: M13-21.  
 Location/Qualifiers  
 source 1..428  
 /organism="Saccharomyces cerevisiae"  
 /db\_xref="X2180-1A"  
 /db\_xref="taxon:4932"  
 /clone\_lib="S. cerevisiae strain X2180-1A"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI"

BASE COUNT 106 a 93 c 72 g 157 t

ORIGIN

Query Match 100.0%; Score 11; DB 20; Length 428;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTAGT 11  
 |||||  
 DB 119 CTGTTCTAGT 129

RESULT 9

T36417 346 bp mRNA EST 11-JAN-1995  
 LOCUS T36417  
 DEFINITION EST101348 S. cerevisiae strain X2180-1A Saccharomyces cerevisiae  
 CDNA 3' end, mRNA sequence.  
 ACCESSION T36417  
 NID 9620234  
 VERSION T36417.1 GI:620234  
 KEYWORDS EST.  
 SOURCE Baker's yeast.  
 ORGANISM Saccharomyces cerevisiae  
 Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 Saccharomycetaceae; Saccharomyces.  
 1 (bases 1 to 346)  
 Weinstock,K.  
 Saccharomyces cerevisiae CDNAs  
 Unpublished (1995)  
 Other ESTs: TC2  
 Contact: Weinstock,K. and Venter,J.C.  
 The Institute for Genomic Research  
 932 Clopper Rd, Gathersburg, MD 20878  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: tdbinfo@tdb.tigr.org  
 For clone availability please contact the TIGR Database  
 (tdbinfo@tdb.tigr.org)

FEATURES  
source

Seq primer: M13-21.  
Location/Qualifiers  
1. .346  
/organism="Saccharomyces cerevisiae"  
/db\_xref="taxon:4932"  
/clone\_lib="S. cerevisiae strain X2180-1A"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT  
81 a 68 c 63 g 130 t 4 others

ORIGIN  
1 CTGTTCTAGT 11  
|||||  
119 CTGTTCTAGT 129

Query Match  
Best Local Similarity 100.0%; Score 11; DB 20; Length 346;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 10  
T48192 328 bp mRNA EST 01-FEB-1995  
LOCUS yb25h05.r1 StrataGene fetal spleen (#937205) Homo sapiens cDNA  
DEFINITION  
Accession T48192  
NID 9650172  
VERSION T48192.1 GI:650172  
KEYWORDS EST  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 328)  
Hiller, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,  
Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Roehling, T., Scheinberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevasks, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478

TITLE  
JOURNAL  
MEDLINE  
COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 221  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: M13RP1  
High quality sequence stop: 221.  
Location/Qualifiers  
1. .328  
/organism="Homo sapiens"  
/db\_xref="GDB:493914"  
/db\_xref="taxon:9606"  
/map\_x: 2; 726H05; 12"  
/clone="IMAGE:72249"  
/clone\_lib="StrataGene fetal spleen (#937205)"  
/tissue\_type="fetal spleen"  
/dev\_stage="fetal"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: spleen; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT. Pooled spleens. Average insert size: 1.0 Kb;

FEATURES  
source

Un1-zAP XR Vector: ~5' adaptor sequence: 5' GAATTCGACGACG  
3' ~3' adaptor sequence: 5' CTCAGATTTTCTTTTCTTTT 3'

BASE COUNT  
71 a 77 c 103 g 71 t 6 others

ORIGIN  
1 CTGTTCTAGT 11  
|||||  
Db 182 CTGTTCTAGT 172

Query Match  
Best Local Similarity 100.0%; Score 11; DB 20; Length 328;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11  
T57378 294 bp mRNA EST 08-FEB-1995  
LOCUS yb51a04.r1 StrataGene fetal spleen (#937205) Homo sapiens cDNA  
DEFINITION  
Accession T57378  
NID 9659239  
VERSION T57378.1 GI:659239  
KEYWORDS EST  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 294)  
Hiller, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,  
Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Roehling, T., Scheinberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevasks, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478

TITLE  
JOURNAL  
MEDLINE  
COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 704  
High quality sequence stops: 216  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the IMAGE  
Consortium (info@image.llnl.gov) for further information.  
Insert Length: 704 Std Error: 0.00  
Seq primer: M13RP1  
High quality sequence stop: 216.  
Location/Qualifiers  
1. .294  
/organism="Homo sapiens"  
/db\_xref="GDB:496335"  
/db\_xref="taxon:9606"  
/clone="IMAGE:74670"  
/clone\_lib="StrataGene fetal spleen (#937205)"  
/tissue\_type="fetal spleen"  
/dev\_stage="fetal"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: spleen; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT. Pooled spleens. Average insert size: 1.0 Kb;  
Un1-zAP XR Vector: ~5' adaptor sequence: 5' GAATTCGACGACG  
3' ~3' adaptor sequence: 5' CTCAGATTTTCTTTTCTTTT 3'

BASE COUNT  
79 a 65 c 60 g 84 t 6 others

ORIGIN  
Query Match  
100.0%; Score 11; DB 20; Length 294;

Best Local Similarity 100.0%; Pred. No. 2.8e+03;  
Matches 11: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTAGT 11  
|||||  
Db 200 CTGTTCTAGT 210

## RESULT 12

T57423 418 bp mRNA EST 08-FEB-1995  
LOCUS T57423/c  
DEFINITION yb51e06.r1 Stratagene fetal spleen (#937205) Homo sapiens CDNA  
clone IMAGE:74722.5, similar to similar to gb:U04217  
PORPHOBILINOGEN DEAMINASE (HUMAN), mRNA sequence.

ACCESSION T57423  
NID 9659284  
VERSION T57423.1 GI:659284  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapelli, B.,  
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rikkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
and Marra, M.

TITLE Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL Genome Res. 6 (9), 807-828 (1996)  
MEDLINE 97044478  
COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu

Insert Size: 1707  
High quality sequence stops: 291 Source: IMAGE Consortium, LNL. This  
clone is available royalty-free through LNL; contact the IMAGE  
Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1707 Std Error: 0.00  
Seq primer: M13RP1  
High quality sequence stop: 291.

## FEATURES

source 1. 418  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="GDB:496387"  
/db\_xref="taxon:9606"  
/clone="IMAGE:74722"  
/clone\_lib="Stratagene fetal spleen (#937205)"  
/tissue\_type="fetal spleen"  
/dev\_stage="fetal"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: spleen; Vector: Bluescript SK-; site\_1:  
EcoRI; site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT. Pooled spleens. Average insert size: 1.0 kb;  
Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGCGACGAG  
3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'"

BASE COUNT 98 a 95 c 112 g 105 t 8 others  
ORIGIN

Query Match 100.0%; Score 11; DB 20; Length 418;  
Best Local Similarity 100.0%; Pred. No. 2.7e+03;  
Matches 11: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTAGT 11  
|||||  
Db 205 CTGTTCTAGT 195

## RESULT 13

T57520 513 bp mRNA EST 08-FEB-1995  
LOCUS T57520  
DEFINITION yb58b01.s1 Stratagene ovary (#937217) Homo sapiens CDNA clone  
IMAGE:75337.3, similar to contains Alu repetitive element, mRNA  
sequence.

ACCESSION T57520  
NID 9659381  
VERSION T57520.1 GI:659381  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapelli, B.,  
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rikkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
and Marra, M.

TITLE Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL Genome Res. 6 (9), 807-828 (1996)  
MEDLINE 97044478  
COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu

Insert Size: 1280  
High quality sequence stops: 321 Source: IMAGE Consortium, LNL. This  
clone is available royalty-free through LNL; contact the IMAGE  
Consortium (info@image.llnl.gov) for further information. Possible  
reversed clone: polyT not found  
Insert Length: 1280 Std Error: 0.00  
Seq primer: -21m13  
High quality sequence stop: 321.

## FEATURES

source 1. 513  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="GDB:505066"  
/db\_xref="taxon:9606"  
/clone="IMAGE:75337"  
/clone\_lib="Stratagene ovary (#937217)"  
/sex="female"  
/dev\_stage="49 year old"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: ovary; Vector: Bluescript SK-; site\_1:  
EcoRI; site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT. Total ovary tissue, normal, caucasian. Average insert  
size: 0.8 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5'  
GAATTCGCGACGAG 3' -3' adaptor sequence: 5'  
CTCGAGTTTCTTTTCTTTT 3'"

BASE COUNT 114 a 104 c 112 g 180 t 3 others  
ORIGIN

Query Match 100.0%; Score 11; DB 20; Length 513;  
Best Local Similarity 100.0%; Pred. No. 2.6e+03;  
Matches 11: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTAGT 11  
|||||  
Db 123 CTGTTCTAGT 133

RESULT 14  
235023/c

LOCUS Z35023 383 bp mRNA EST 18-JUL-1994  
 DEFINITION AT35362 AC16H Arabidopsis thaliana cDNA clone TA082.3', mRNA  
 SEQUENCE  
 ACCESSION Z35023  
 NID 9510771  
 VERSION Z35023.1 GI:510771  
 KEYWORDS EST.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 Arabidopsis.  
 REFERENCE 1 (bases 1 to 383)  
 CNRS.  
 TITLE The Arabidopsis thaliana transcribed genome: the GDR cDNA program  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Berthomieu P., Guerrier D., Giraudat J.  
 Genetique Moleculaire d'Arabidopsis  
 ISV - UPR40, CNRS  
 Avenue de la Terrasse, 1198 Gif-sur-Yvette Cedex, France  
 Email: Giraudat@cnrs-gif.fr.  
 FEATURES  
 SOURCE Location/Qualifiers  
 1..383  
 /organism="Arabidopsis thaliana"  
 /strain="ecotype Columbia"  
 /db\_xref="taxon:3702"  
 /clone="TA082"  
 /clone\_1lb="AC16H"  
 /note="Vector: Lambda ZAPIT; tissue-cell suspension  
 culture of ecotype Columbia; clone\_library=AC16H; Cloning  
 vector: Lambda ZAPIT; Physiological condition: cycling  
 cells."  
 BASE COUNT 106 a 82 c 69 g 121 t 5 others  
 ORIGIN  
 Query Match 100.0%; Score 11; DB 20; Length 383;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTGTTTCAGT 11  
 |||||||||  
 Db 129 CTGTTTCAGT 119  
 U1T 15  
 LOCUS 238728 295 bp mRNA EST 31-OCT-1994  
 DEFINITION HSCONB082 normalized infant brain cDNA Homo sapiens cDNA clone  
 c-0nb08.3', mRNA sequence.  
 ACCESSION 238728  
 NID 9561120  
 VERSION 238728.1 GI:561120  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 295)  
 Aufrey, C., Behar, G., Bois, F., Bouchier, C., da Silva, C.,  
 Devignes, M.D., Duprat, S., Houligatte, R., Juneau, M.N., Lamy, B.,  
 Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, Y.,  
 Sebastiani-Kabackis, C. and Tessier, A.  
 IMAGE: molecular integration of the analysis of the human genome  
 and its expression  
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
 JOURNAL 95277534  
 MEDLINE  
 COMMENT Contact: Genethon  
 Genexpress-Genethon  
 Genethon Centre de recherche sur le Genome Humain

1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE  
 Tel: 33169472800  
 Fax: 33160778698  
 Email: genexpress@genethon.fr  
 Single read. 18 T removed at sequence 5' end  
 Genexpress library Idt: C; Genexpress\_sequence\_idt: alc-0nb08  
 Seq primer: (-21)M3-universal.  
 FEATURES  
 SOURCE Location/Qualifiers  
 1..295  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="c-0nb08"  
 /clone\_1lb="normalized infant brain cDNA"  
 /sex="Female"  
 /tissue\_type="total brain"  
 /dev\_stage="3 months old"  
 /note="Organ: brain; Vector: lambda BA; Site\_1: HindIII;  
 Site\_2: NotI; sex=Female; dev\_stage=3 months old;  
 isolate=muscular atrophy patient; tissue\_type=total  
 brain; total mRNA was oligo-(dT) primed and directionally  
 cloned 5' -> 3' into the HindIII -> NotI sites of the  
 lambda BA vector. Clone library from B. Soares, Psychiatry  
 Dept. Columbia University, USA  
 Bento Soares, P.N.A.S. in press  
 Normalization\_method:  
 BASE COUNT 81 a 60 c 62 g 92 t  
 ORIGIN  
 Query Match 100.0%; Score 11; DB 20; Length 295;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTGTTTCAGT 11  
 |||||||||  
 Db 23 CTGTTTCAGT 33

Search completed: September 17, 1999, 21:28:00  
 Job time: 14264 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:01:12 ; Search time 1962.17 Seconds

(without alignments)  
16.208 Million cell updates/sec

Title: US-08-956-518a-89

Perfect score: 10

Sequence: 1 ACCCACACAG 10

Scoring table: IDENTITY\_NUC

Searched: 679419 segs, 1590154680 residues

abase :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_cm: \*  
4: gb\_cv: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_pl1: \*  
8: gb\_pl2: \*  
9: gb\_pl3: \*  
10: gb\_pr2: \*  
11: gb\_pr3: \*  
12: gb\_ro: \*  
13: gb\_st: \*  
14: gb\_sts: \*  
15: gb\_sy: \*  
16: gb\_un: \*  
17: gb\_v1: \*  
18: em\_fun: \*  
19: em\_htg: \*  
20: em\_hum1: \*  
21: em\_hum2: \*  
22: em\_in: \*  
23: em\_om: \*  
24: em\_or: \*  
25: em\_ov: \*  
26: em\_pat: \*  
27: em\_ph: \*  
28: em\_pl: \*  
29: em\_ro: \*  
30: em\_sts: \*  
31: em\_sy: \*  
32: em\_un: \*  
33: em\_v1: \*  
34: gb\_hg1: \*  
35: gb\_hg2: \*  
36: gb\_in1: \*  
37: gb\_in2: \*  
38: em\_ba1: \*  
39: em\_ba2: \*  
40: em\_hum3: \*  
41: em\_hum4: \*  
42: gb\_pr4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No. 10 100.0 12328 1 AE000815

Query Match Length DB ID

SUMMARIES

Description AE000815 Methanoba

C	2	10	100.0	18988	1	AE000876	AE000876 Methanoba
C	3	10	100.0	2280	1	ANNOXER	Y13471 Anaerobacillus n
C	4	10	100.0	257000	1	AP000002	AP000002 Pyrococcus
C	5	10	100.0	1417	1	ASRRNA16S	X81063 Actinomyces
C	6	10	100.0	4638	1	ATPRIRP	X04833 A. rhizogen
C	7	10	100.0	1928	1	BACMCASE	D01057 B. subtilis
C	8	10	100.0	26170	1	BC17ODEGR	Z73234 B. subtilis
C	9	10	100.0	2435	1	BS14GLUC	X67044 B. subtilis
C	10	10	100.0	1800	1	BSEGLUC2	X04689 Bacillus su
C	11	10	100.0	233780	1	BSEGLG	Z29076 B. subtilis
C	12	10	100.0	217420	1	BSUB0010	Z99113 Bacillus su
C	13	10	100.0	3483	1	BSUB0017	Z99120 Bacillus su
C	14	10	100.0	485	1	BT004366	U04366 Bacillus th
C	15	10	100.0	1532	1	BUHARGSC	L18930 Buchnera ap
C	16	10	100.0	19650	1	D89332	D89332 Bifidobacte
C	17	10	100.0	18252	1	D90736	D90736 Escherichia
C	18	10	100.0	15676	1	D90837	D90837 Escherichia
C	19	10	100.0	13548	1	D90858	D90858 E. coli geno
C	20	10	100.0	133859	1	D90899	D90899 Synecocyst
C	21	10	100.0	137740	1	D90900	D90900 Synecocyst
C	22	10	100.0	130690	1	D90901	D90901 Synecocyst
C	23	10	100.0	132419	1	D90907	D90907 Synecocyst
C	24	10	100.0	145709	1	D90914	D90914 Synecocyst
C	25	10	100.0	14852	1	EC4HPADNA	Z37980 E. coli hpa
C	26	10	100.0	338534	1	ECOW93	U14003 Escherichia
C	27	10	100.0	4680	1	ECTOR	X73888 E. coli gene
C	28	10	100.0	4235	1	ECU58768	U58768 Escherichia
C	29	10	100.0	690	1	FR130884	AJ130884 Flexispir
C	30	10	100.0	8983	1	H1008875	U08875 Haemophilus
C	31	10	100.0	9221	1	H1008876	U08876 Haemophilus
C	32	10	100.0	1680	1	HMOFRA	X73823 H. mediterr
C	33	10	100.0	36063	1	MLT15184	U15184 Mycobacteri
C	34	10	100.0	38721	1	MSGY219	AD000013 Mycobacte
C	35	10	100.0	42061	1	MSGY223	AD000019 Mycobacte
C	36	10	100.0	34150	1	MTCY190	Z70283 Mycobacteri
C	37	10	100.0	20760	1	MTCY21D4	Z80775 Mycobacteri
C	38	10	100.0	35187	1	MTCY441	Z80225 Mycobacteri
C	39	10	100.0	2420	1	MTHDRBC	X81134 M. thermoa
C	40	10	100.0	13514	1	MTRFVIA	X68366 M. thermof
C	41	10	100.0	68848	1	MTV043	AL022004 Mycoplasma
C	42	10	100.0	420	1	MYCTGG	L15242 Mycoplasma
C	43	10	100.0	694	1	PAL239040	AJ239040 Prevotell
C	44	10	100.0	193772	42	AC007073	AC007073 Homo sapi

#### ALIGNMENTS

RESULT 1  
AE000815  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

TITLE  
Complete genome sequence of Methanobacterium thermoautotrophicum  
deltaH: functional analysis and comparative genomics

JOURNAL MEDLINE  
J. Bacteriol. 179 (22), 7135-7155 (1997)  
98037514  
2 (bases 1 to 12328)  
REFERENCE  
AUTHORS  
Smith, D. R.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (10-ANG-1997) Genomics and Technology Development, Genome

FEATURES	Location/Qualifiers
source	1..,12328

repeat\_region

```

/organism="Methanobacterium thermoautotrophicum
/strain="delta H"
/db_xref="taxon:2166"
/clone="MTH"

```

gene  
CDS

gene	70. .603
CDS	/gene="MTH293" 70. .603

```

/ gene="MT1293"
/ note="Function Code:14_00 - Unknown, ; similar to,
sp.LN:YAW5_SCHPO, p()=0.29, pid=8"
/ codon_start=1
/ transl_table=11
/ product="unknown"
/ protein_id="AAB84799.1"
/ db_xref="PID:g2621346"
/ db_xref="GI:2621346"
/ translation="MIGMSIDDIYMTLEERISSAFRDEETWELRHANPMSVYTLTAL
PLVAALWSEKEMGMAWAIIPVTLVLTWYINLPRLEKPESTDSSASSVAGERVWLR
WLYDPENHRRVPIVILNLIATGTVPIWGYRLEVPVLLGFAMVYIGKLMYLDKRVV
MVDQMEKREVEYRDMWY"
/ complement(747...1211)
/ ...

```

```

/product="DNA-dependent RNA polymerase, subunit A' "
/protein_id="AAB84803.1"
/db_xref="PID:92621350"
/db_xref="GI:2621350"
/translation="MRGLKKIGIEFGMLSPEDIRKISVAQIVPTPYNEDGYPIEN
GLMDPLGLVDIRPERCWTGARGCEGCGHFSILARPLVHVGFADTIHKILRSTCKR
CGAVLLGTETEEYIRGLKPMWRKRKTSRH"
3825 4094
/gene="mtr298"
/contig="mtr298"
/cds_start="3825"
/cds_end="4094"

```

gene	complement(747. .1211)
CDS	/gene="MTH294" complement(747. .1211)

gene  
CDS

```

/gene="MTH294"
/note="Function Code:14.00 - Unknown, ; similar to,
sp.LIN:RS27_ENTH1 AC:P38654, p(-0.64, pId=088"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB84800.1"
/db_xref="Pid:g2621347"
/db_xref="GI:2621347"
/translation="MQGIFDSGGISVSKLIKILIVCLVLFGLASFNAGMLNTEVDS
QNNFSYDVNTIRKNPITSSNNESIDIRIPMEENWVWDSRCPCGAPFSYSDVPEYEDD
GIWYEKRYREYECGILFMGTDVYRIYLMBEAAEQIRKEAMKRGKPPKAI"
1459..1803
/gene="MTH295"
1459..1803

```

```

/product="DNA-dependent RNA polymerase, subunit A' "
/protein_id="AAB84804.1"
/db_xref="FID:g2621351"
/db_xref="GI:2621351"
/translation="MEKEENPTLVNRYSLARDRCPHCEEEOEIKLPKPIVEGG
DYKLTPEVKEKRLERISDDDALIGVNPQVAPREMTVLTPVP"
4097. 6091
/gene="MTH29"
4097. 6091
/gene="MTH29"
CDS
/gene="MTH29"

```

```

/gene="MTH295"
/note="Function Code:14.00 - Unknown, ; similar to,
gp:GI:91839544 UN:S82218, p():0.91, pId=07%"
/codon_start=1
/translation_table=11
/product="unknown"
/protein_id="AAB84801.1"
/db_xref="PID:g2621348"
/db_xref="GI:2621348"
/translation="MCIVSSWVGQPMIMAGILEEYVHKSGDPSISDPTLNKRPAYD
YSDVMDGSDGIDSARPPGAAPGHRIRGSPVYLLYDDELTGHWAVYYDVDAQGYSEYD
SMDPNAGATMIN"
2168..3427
/gene="MTH296"
2168..3427

```

/note="Function Code:10.02 - Metabolism of Macromolecules,  
 Transcription--mRNA synthesis and modification (includes  
 regulators) ; similar to, sp:LN:RPAL-METH AC:F09846,  
 P(1):4.4E-250, Pld=1318"  
 /codon\_start=1  
 /transl\_table=1  
 /product="DNA-dependent RNA polymerase, subunit A' "  
 /protein\_id="AAB84805.1"  
 /db\_xref="PDB:92621352"  
 /db\_xref="GI:2621352"  
 /translation="MPSITLKNGRSESDITLKVLDILRINRKNMEGAGAPOLIV  
 EDIMELIQAHTTQEDNDEASGVTPAROSGOLLTLMORLKSQKGSSELSKSPFG  
 NLSGRNFSAKRTYISPDPNISINEGVPELIAEVTVPYVETMNDIRAREIYENQ  
 DVHGAGNATVPDGRIRISCKPTAEIENKPYVVERHLKDDIVLENNQSLHML  
 SLMAHRAVRYEYKTKFLNLCVCPYPNADFDQDENIVHLOTTEESRAAKTLMQVQEHM

```

/ gene="MTH296"
/ note="Function Code:14.00 - Unknown, i similar to,
sp.LN:KC2B.CAEEL.AC:P28548, p()=0.9, pid=088"
/ codon_start=1
/ transi_table=11
/ product="unknown"
/ protein_id="AAB84802.1"
/ db_xref="PID:g2621349"
/ db_xref="GI:2621349"
/ translation="MWQPLRGSEGEDTRPPQAPRNKAAQPRITLTASDQENLGLTDE

```

gene  
/gene="MTH300"  
6188..6328  
DSD 6188..6328

/gene="MTH300"  
 /note="Function Code:14.00 - Unknown, ; similar to,  
 gp:GI:9642348, p()=0.51, pld=14%"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="AAB84806.1"  
 /db\_xref="PID:g2621353"  
 /db\_xref="GI:2621353"  
 /translation="MFRKDPQLMKPLPSLPLOCFOLFSGYHPGNATSGFDGSSAD  
 IN"  
 6367. .6885  
 /gene="MTH301"  
 6367. .6885  
 /gene="MTH301"  
 /note="Function Code:14.00 - Unknown, ; similar to,  
 sp:NN:SCOL CAEEL AC:009450, p()=0.98, pld=13%"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="AAB84807.1"  
 /db\_xref="PID:g2621354"  
 /db\_xref="GI:2621354"  
 /translation="MRRALIMAMICSTLAISGCTGTGNTLTSESTGYPAEVEEY  
 RVASGCTIDSEYKGLNLSNOCILIFARIVLDSNDKNIIVREYGNESFDGL  
 SSALRLPEYVDMAGRTYRMWDSWKPNHVGIGEGAGDLSPLLRAVRVILTEYR  
 STGLPDRVTP"  
 complement(6975. .7520)  
 /gene="MTH302"  
 complement(6975. .7520)  
 /gene="MTH302"  
 /note="Function Code:14.00 - Unknown, ; similar to,  
 100.0%; Score 10; DB 1; Length 12328;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAG 10  
 |||||  
 DB 3359 ACCCACACAG 3368

RESULT 2  
 AEO00876 18988 bp DNA BCT 16-APR-1998  
 LOCUS  
 DEFINITION Methanobacterium thermoautotrophicum from bases 947295 to 966282  
 (section 82 of 148) of the complete genome.  
 AEO00876 AEO00666  
 g2622140  
 AEO00876.1 GI:2622140

J  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

Methanobacterium thermoautotrophicum.  
 Methanobacterium thermoautotrophicum  
 Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
 Methanobacterium.  
 1 (bases 1 to 18988)  
 Smith,D.R., Doucette-Stamm,L.A., Deloughery,C., Lee,H.-M.,  
 Dubois,J., Adrege,T., Bashirzadeh,R., Blakely,D., Cook,R.,  
 Gilbert,K., Harrison,D., Hoang,L., Keagle,P., Lumu,W., Pothier,B.,  
 Qiu,D., Spadofora,R., Vicare,R., Wang,Y., Wierzbowski,J.,  
 Gibson,S., Jiwani,N., Garuso,A., Bush,D., Sater,H., Patwell,D.,  
 Prabakar,S., McDougall,S., Shimer,G., Goyal,A., Pletcovski,S.,  
 Church,G.M., Daniels,C.J., Mao,J.-I., Rice,P., Nolling,J. and  
 Reeve,J.N.  
 Complete genome sequence of Methanobacterium thermoautotrophicum  
 deltaH: functional analysis and comparative genomics  
 J. Bacteriol. 179 (22), 7135-7155 (1997)  
 98037514  
 2 (bases 1 to 18988)  
 Smith,D.R.  
 Direct Submission  
 Submitted (10-Aug-1997) Genomics and Technology Development, Genome  
 Therapeutics Corporation, 100 Beaver Street, Waltham, MA

02154-8448, USA

FEATURES

Source

Location/Qualifiers

1. .18988

/organism="Methanobacterium thermoautotrophicum"

/strain="delta H"

/db\_xref="taxon:2166"

/clone="MTH"

complement(87. .1172)

/gene="MTH1037"

complement(87. .1172)

/gene="MTH1037"

/note="Function Code:14.00 - Unknown, ; similar to, sp:LN:P044\_RAT AC:P38718, p(-)=0.75, pid=068"

/codon\_start=1

/transl\_table=11

/product="unknown"

/protein\_id="AAB85533.1"

/db\_xref="PID:g2622141"

/db\_xref="GI:2622141"

/translation="MLIAQNQLQFILEAVI IHVGIILFNVAAPISLVFLGTVLT  
VALAIFSDLAAPFLPFLSHHEPTHEGFAVLFNTVAASNLTEAGIGSAVKK  
LSILFEVIAISGGIMHRSFLVILGMARGLIMSPFRSTITKKSYSFLAKR  
GARMEELSKVINKSVLSPMLRTREEMNPBLSYLKNTPTMGVQSGCYKSKC  
LGGADGITTPITVYIQLGLPHYITFCVLYVKKDIIDYMLDGIHVAVFDAGFELLE  
LFWMVITFCGLSVLRRYQORLNGSRMYIGREALLIGSLAFLSOSIMGLFIPIRSE  
NSSALTYIIISILVMAHTVYIKRTIS"

complement(1335. .2204)

/gene="MTH1038"

complement(1335. .2204)

/gene="MTH1038"

/note="Function Code:14.01 - Unknown, Conserved protein;  
similar to, sp:LN:YPUA\_BACSU AC:P31847:P37951,  
p(-)=4.7E-20, pid=238"

/codon\_start=1

/transl\_table=11

/product="conserved protein"

/protein\_id="AAB85534.1"

/db\_xref="PID:g2622142"

/db\_xref="GI:2622142"

/translation="MYMROIPLPLIAFLVAPVFAVSGFVYTLGRTSSNPYSKSA  
WMDYFKAKTKRDLQSDADIKIVTASVENVESRGYTGAVYSPSOLSCAMVDLSISDGIR  
VSDYTKIRVTPDEMVASALRSSGIDGYYVAVSPVASEALAGVLSKAEIGAVGQ  
IPEPAKRVSEIYILOSRLVYENATGDRVAELFDEYKNRTOSQLDDPADIDQHVVD  
VSOOMNINLETGTQOVADSVAAAPVQSGSLTEFKRLREGVSQOVGSGILDIYAFEL  
QSILYNYIMGTASP"

complement(2270. .3187)

/gene="MTH1039"

complement(2270. .3187)

/gene="MTH1039"

/note="Function Code:14.01 - Unknown, Conserved protein;  
similar to, sp:LN:Y257\_METRA AC:Q57705, p(-)=1.7E-54,  
pid=388"

/codon\_start=1

/transl\_table=11

/product="conserved protein"

/protein\_id="AAB85535.1"

/db\_xref="PID:g2622143"

/db\_xref="GI:2622143"

/translation="MAWTLDESQRKRMGRYFVGEHGHSAVYTCMTKTSYINEGYC  
YKRKFGIRSHRCLQMSPVFCQOKLFCWRDLSTGTWMEGHPEHDAIDGALIPA  
KRIILCGYLNERADKIKVMESODPTAAISLAGEPMILYPMOGLLREFHRRNFTFL  
VTNGLAPMNLKESEPTOLYISLDADRDYKELCRPQIGAMVDLNSLELMPFS  
CRVYLRITAVRHINKPDEGFARMEARPDFEVKAYMYIGYSRRRLDIENPLPFE  
VHFEAEELASAEMEYDESRERVILLA"

3831. .4487

/gene="MTH1040"

3831. .4487

/gene="MTH1040"

/note="Function Code:14.00 - Unknown"

/codon\_start=1

/transl\_table=11

/product="unknown"

/protein\_id="AAB85536.1"

```

/db_xref="pid:92622144"
/db_xref="GI:2622144"
/translation="MSTIPGOLFPODSSGDKITAKSTSLPKVDAAKKYKRYKRYKRY
HASKGYAKITRTKTKYRTAAKTKRKKYRSATYSTSRKAYRAISDEL
NNLODEBELKIASINNRLNHRSGPTAGVKGDCWGLSDMSARKLANGK
VKVVOGATSSSSRRHMLHVYSEGRWTSPELVTMKRYSKRYTATCGRATRVATYNN
"
4683..5372
/gene="MTH1041"
/feature="CDS"
4683..5372
/gene="MTH1041"
/note="Function Code:1.01 - Carbohydrate Metabolism,
glycolysis--gluconeogenesis ; similar to, sp:LN:RITS_METUA
AC:058923, p(-)1E-63, pld-51%"
/codon_start=1
/transl_table=11
/product="triosephosphate isomerase"
/protein_id="AAB85537.1"
/db_xref="PID:92622145"
/db_xref="GI:2622145"
/translation="MLEDLEKDPPIYILNFKYLESTGERALELASICGDVADETV
NMAYAPQMDLHRVSDAVEIPVLAQHIDAVDAGHTGSIIECARBDAGAAGTILNHE
KRMQLADIEMWISMKLEMMSVVCTNNMTTAAALGPDFAVEPELIGSIPVS
RAPEVITGSYDAVKKNPEYVLCGAGISTGDDMKAAVLDGAEGVLLAGIILADSP
RDALDLVSKY"
5369..6604
/gene="MTH1042"
/feature="CDS"
5369..6604
/gene="MTH1042"
/note="Function Code:1.01 - Carbohydrate Metabolism,
glycolysis--gluconeogenesis ; similar to, sp:LN:PK_METTE
AC:120971, p(-)2.3E-137, pld-62%"
/codon_start=1
/transl_table=11
/product="3-phosphoglycerate kinase"
/protein_id="AAB85538.1"
/db_xref="PID:92622146"
/db_xref="GI:2622146"
/translation="MMYIPVSFKRTMDIEVTKYLVAVDINSVPDNDGILTDT
RMLHATTIRELSDRKATYVMAHQSPGKNDPTTLEQHRALVSGILRPVKVEDIF
GSAAREKISGLGGEITLLENVRFSESEVLRKDEQAEHLVRLKAPLDIVDINAF
AAAHRSQPSLVGFALRVPSAAGVMEERELTLOGALENRPVYVYGKVDISLV
MKNVLENGSADLVLTGLVANIFLAGGVKIGVNMDEFISRGCDFFIVAKLKRFR
PERIIVPVDAVCRDGRVDVPVKIIPNHPIODIGMETITVLRIRREARLPANPA
GVFENPESICTEDILNATISSSGFSLIGGHLAAAVAKMGFEDSINHSSGGGASIS
LAGEELPARYVLEESNHP"
Complement(6645..6718)
/note="74 bp inverted repeat includes part of MTH1043
(tRNA-Thr anticodon:ugu); 84% ID to interval
678954-679026"
6645..6718
/gene="MTH1043"
/note="codon recognized: ACA"
/product="tRNA-Thr"
6645..6718
/gene="MTH1043"
/complement(6651..6711)
/note="61 bp direct repeat includes part of MTH1043
(tRNA-thr anticodon:ugu); 85% ID to interval
954496-954436"
6726..6800
/gene="MTH1044"
/note="codon recognized: CCA"
/product="tRNA-Pro"
6726..6800
/gene="MTH1044"
/feature="CDS"
6855..6927
/gene="MTH1045"
/note="codon recognized: TAC"
/product="tRNA-Tyr"
6855..6927
/gene="MTH1045"
/feature="CDS"
7056..7127

```

```

/gene="MTH1046"
/note="codon recognized: CAC"
/product="tRNA-Asp"
7056..7127
/gene="MTH1046"
/feature="CDS"
7136..7209
/gene="MTH1047"
/note="codon recognized: AAA"
/product="tRNA-Lys"
7136..7209
/gene="MTH1047"
/complement(7142..7202)
/note="61 bp direct repeat includes part of MTH1047
(tRNA-ys anticodon:uuu); 85% ID to interval 954005-953945"
7252..7485
/gene="MTH1048"
/feature="CDS"
7252..7485
/gene="MTH1048"
/note="Function Code:10.02 - Metabolism of Macromolecules,
transcription--mRNA synthesis and modification (includes
regulators) ; similar to, sp:LN:RPOH_SULAC AC:P11521,
p(-)1.6E-17, pld=56%"
/codon_start=1
/transl_table=11
/product="DNA-dependent RNA polymerase, subunit H"

Query Match      100.0%; Score 10; DB 1; Length 18988;
Best Local Similarity 100.0%; Pred No. 2.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 ACCCACACAG 10
Db      2660 ACCCACACAG 2651

RESULT      3
ANNOXEF/c    ANNOXEF      2280 bp      DNA      BCT      01-DEC-1998
LOCUS      Anacystis nidulans hoxE, ORF1 genes.
DEFINITION  Y13471
ACCESSION  Y13471
VERSION    93947769
KEYWORDS   hoxE gene; hoxF gene; ORF1.
SOURCE     Synecchococcus PCC6301.
ORGANISM   Bacteria; Cyanobacteria; Chroococcales; Synecchococcus.
REFERENCE  1 (bases 1 to 2280)
AUTHORS    Bolson,G., Schmitz,O., Schmitz,B. and Bothe,H.
TITLE      Unusual gene arrangement of the bidirectional hydrogenase and
           functional analysis of its diaphorase subunit Hoxu in respiration
           of the unicellular cyanobacterium anacystis nidulans
           Curr. Microbiol. 36 (5), 253-258 (1998)
JOURNAL    2 (bases 1 to 2280)
MEDLINE    Bolson,G.
REFERENCE  Direct Submission
           Submitted (29-MAY-1997) G. Bolson, Universitaet zu Koeln,
           Botanisches Institut, Gyrhofstr. 15, D-50931 Koeln, FRG
FEATURES
Source     1..2280
           /organism="Synecchococcus PCC6301"
           /strain="1402-1"
           /db_xref="taxon:1139"
           /clone_lib="lambdaGEM-11"
           53..>2280
           /gene="hoxE"
           /note="putative"
           53..2280
           /gene="hoxE"
           74..571
           /gene="hoxE"
           /codon_start=1
           /transl_table=11

```



Query Match	100.0%	Score 10;	DB 1;	Length 2280;
100.0%	100.0%	Prod. No. 3.1e+03;		
10; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1 ACCCACACAG 10				
Db 463 ACCCACACAG 454				
RESULT 4				
LOCUS AP000002				
DEFINITION AP000002 257000 bp DNA BCT 08-FEB-1999				
ACCESSION AB0009475 AB009476 AB009477 AB009478 AB009479 AB009480				
AB009481 AB009482 AB009483 AB009484				
AB009485				
AB009486				
AB009487				
AB009488				
AB009489				
AB009490				
AB009491				
AB009492				
AB009493				
AB009494				
AB009495				
AB009496				
AB009497				
AB009498				
AB009499				
AB009500				
AB009501				
AB009502				
AB009503				
AB009504				
AB009505				
AB009506				
AB009507				
AB009508				
AB009509				
AB009510				
AB009511				
AB009512				
AB009513				
AB009514				
AB009515				
AB009516				
AB009517				
AB009518				
AB009519				
AB009520				
AB009521				
AB009522				
AB009523				
AB009524				
AB009525				
AB009526				
AB009527				
AB009528				
AB009529				
AB009530				
AB009531				
AB009532				
AB009533				
AB009534				
AB009535				
AB009536				
AB009537				
AB009538				
AB009539				
AB009540				
AB009541				
AB009542				
AB009543				
AB009544				
AB009545				
AB009546				
AB009547				
AB009548				
AB009549				
AB009550				
AB009551				
AB009552				
AB009553				
AB009554				
AB009555				
AB009556				
AB009557				
AB009558				
AB009559				
AB009560				
AB009561				
AB009562				
AB009563				
AB009564				
AB009565				
AB009566				
AB009567				
AB009568				
AB009569				
AB009570				
AB009571				
AB009572				
AB009573	</			

REFERENCE AUTHORS

Kawarabayashi, National Institute of Technology and Evaluation, Biotechnology Center, 2-chome 49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan (E-mail: genome@nitech.go.jp, Tel: +81-3-3481-8951, Fax: +81-3-3481-8424)

2 (sites)

Kawarabayashi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otsuka, K., Nakazawa, H., Takamiya, M., Ohikuni, Y., Funahashi, T., Tanaka, T., Kudo, Y., Yamazaki, J., Kushioka, N., Ouguchi, A., Aoki, K., Nakamura, Y., Robb, T. F., Horikoshi, K., Masuchi, Y., Shikuya, H. and Kikuchi, H.

Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3

DNA Res. 5 (2), 55-76 (1998)

98344137

JOURNAL MEDLINE COMMENT

On Jul 28, 1998 this sequence version replaced g1:3130553

g1:3130577 g1:3130627 g1:3130656 g1:3130685 g1:3130705 g1:3130737

g1:3130785 g1:3130824 g1:3130828

AB009475-AB009484: submitted (10-DEC-1997)

Kawarabayashi, Y. is officially affiliated with the National Institute of Bioscience and Human Technology, Tsukuba, Ibaraki 305-0046, Japan.

Robb, T. F. is at the Center of Marine Biotechnology, University of Maryland, Baltimore, MD, USA.

Horikoshi, K. is at the Japan Marine Science and Technology Center, Yokosuka, Kanagawa 237-0061, Japan.

Masuchi, Y. is at the University of Tokyo, Meguro, Tokyo 153-0041, Japan.

Shizuya, H. is at the California Institute of Technology, Pasadena, CA, USA.

The other authors are at the National Institute of Technology and Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan.

All the sequence with length 100 codons or more between ATG or GTG and stop codon are defined as CDS

Homology analysis is performed by Smith-Waterman algorithm against GenBank and Genpept release 103; EMBL release 52.0; SwissProt release 34.0; PIR-Protein release 54.0; and OML release 29.5.

E-mail address for comments and questions: genome@nitech.go.jp

Restriction map, ORF organization, sequence alignment and more information are available at W.W.W. site of Biotechnology Center, URL: <http://www.bio.nitech.go.jp/>.

Location/Qualifiers

1. 257000

/organism="Pyrococcus horikoshii"

/strain="OT3"

/db\_xref="taxon:53953"

/map="287001-544000 nt. position"

1363. 2664

/gene="PH0318"

1363. 2664

/gene="PH0318"

/note="similar to PIR:564485 percent identity: 55.181 in 403aa; PIR:574842 percent identity: 49.869 in 397aa; SwissProt:P33221 percent identity: 50.773 in 404aa"

/codon\_start=1

/transl\_table=11

/product="433aa long hypothetical phosphoribosylglycinamide formyl transferase"

/protein\_id="BA829382.1"

/db\_xref="PID:d1030335"

/db\_xref="PID:g3256709"

/db\_xref="GI:3256709"

/translation="NYVMYIKRDELGATPDSAKILLGSGELGKEIAEAOALGVEIVAVDYANPAMOVAAHRSYGVGMDKDFKMSYVEREKPAIIPHEIATNLDLFFPEKQIVPPNARATWIMHRRKRLKETLYEKVPTSRMTATIDLELEACQIGTGPCHTKAIMSSGSGSYFVGPEDIIPKAMEAKTKARGSAEKITVEHIDFDEVELAVRHFEDNGEIVTFPFPGVGHYQIDGDYHRSWQPAEISEKAERVYRIAKRIDVGLGGLIFGVMEFGEVYKVALENSPRHDGTGMTLASHPGFSEFALHRAVGLPIPGEMVQYRLPDMILIPATVHTIKAVGSGSPRFGIKALISVPATVATLQPEAYVGRRIALMAWDQVYAKRAEMVAHMETLFRSSDWMHQNTERKHLIR"

complement(1645..1974)

/gene="PH0319"

complement(1645..1974)

CDS

gene

CDS

FEATURES

source

gene

CDS

gene

CDS

complement(1645..1974)

[illegible]

```
FEATURES
    source
        Location/Qualifiers
            1..1417
                /organism="Actinomyces sp."
                /strain="ATCC 49338"
                /db_xref="taxon:29317"
    RNA
        1..1417
            /product="16S ribosomal RNA"
    BASE COUNT
        273 a 339 c 506 g 299 t
    ORIGIN

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 1417;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCCACACAG 10
Db 440 ACCCACACAG 431

JULT 6
APPRIS 4638 bp DNA BCT 26-JUL-1995
LOCUS A. rhizogenes plasmid pR14b replicator region (hairy-root-inducing
DEFINITION plasmid).
ACCESSION X04833
NID X04833.1 GI:39113
VERSION autonomous replication; origin of replication; plasmid; repa gene;
KEYWORDS repB gene; repC gene; R1 plasmid; unidentified reading frame.
SOURCE Agrobacterium rhizogenes.
ORGANISM Agrobacterium rhizogenes.
REFERENCE 1 (bases 1 to 4638)
AUTHORS Nishiyuchi, R., Takamami, M. and Oka, A.
TITLE Characterization and sequence determination of the replicator
JOURNAL Mol. Genet. 206, 1-8 (1987)
COMMENT Data kindly reviewed (23-SEP-1987) by Oka A.
FEATURES
    source
        Location/Qualifiers
            1..4638
                /organism="Agrobacterium rhizogenes"
                /plasmid="pR14b"
                /strain="a transjugant of cured C58 and A. rhizogenes A4"
                /db_xref="taxon:359"
                <1..164
                    /note="put. ORF1 (54 AA) (1 is 2nd base in codon)"
                /codon_start=3
                /transl_table=11
                /protein_id="CAA28526.1"
                /db_xref="PID:g809649"
                /db_xref="GI:809649"
                /db_xref="SPTREMBL:Q44434"
                /translation="MLEVSGRGRHDAFSPRSTLQALLLGQTSPTQTHSLVAIF
                MAISLRHWK
                complement(217..420)
                    /note="put. ORF 4 (AA 1-204)"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA28528.1"
                /db_xref="PID:g380731"
                /db_xref="GI:580731"
                /db_xref="SPTREMBL:Q44435"
                /translation="MLEVSGRGRHDAFSPRSTLQALLLGQTSPTQTHSLVAIF
                AKLEKPRMOLAPILLEGCELLG"
                308..313
                    /note="put. RNA binding site"
                324..1538
                    /note="put. repA (AA 1-404)"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA28527.1"

RBS
    repeat_unit
        /db_xref="PID:g39114"
        /db_xref="GI:39114"
        /db_xref="SWISS-PROT:P05682"
        /translation="MAKSVLKAAPVAVVGTALMERHADALSSLOAHHLKVPPEHSEK
        GTRFPGPSEASKLIGVGESEYLRQTAEMPELNTYMSPGGRMFSIDIVIRKYMQV
        GRGNRYLPHRRGEOLOYISVNMFGSGSKTTPAHLAOYLAMRGYRVLAIDLDPOA
        SLSALGSOPEPTDVGPNELTYGAIIRPDGOVALTERVIRGTYIPDLILIGNLEMEFE
        HDIPRALMRKEEDTLFYGRISSVIDIADNDIVVYIDCPQIGYITLSALTAITSIL
        VYVHPQMLDVMSNOFLAMTSLNLRLEINAGAKFKEMKRYLLITREPSPDQONQNG
        YLRISGENYLNPFMLKTTAVSDAGLTNQLFEVERGLFTSTYDALAMNANVDEI
        EILIKKAWGRPT"
        complement(426..429)
            /note="put. RNA binding site"
        complement(585..941)
            /note="put. ORF 5 (AA 1-119)"
            /codon_start=1
            /transl_table=11
            /protein_id="CAA28529.1"
            /db_xref="PID:g580732"
            /db_xref="GI:580732"
            /db_xref="SPTREMBL:Q44436"
            /translation="MEGVNVSDDSDFCCHLLIIPYSAVERFVRANVRLAPKECR
            KARLRIEINGODISAHSEVLRQMRGGGLTRPTPEIHRHNDKLKLAAGTMQVAVP
            ADLRIILPNHNDIFN"
            complement(947..950)
                /note="put. RNA binding site"
            complement(1510..1620)
                /note="put. ORF 6 (AA 1-36)"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA28531.1"
                /db_xref="PID:g39118"
                /db_xref="GI:39118"
                /db_xref="SPTREMBL:Q44437"
                /translation="MERFPAVLISAGDVSGASVETPRMCFRLMAYPMLE"
                1606..1610
                    /note="put. RNA binding site"
                1619..2537
                    /note="put. rep B (AA 1-312)"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="CAA28530.1"
                    /db_xref="PID:e189998"
                    /db_xref="PID:g1333706"
                    /db_xref="GI:1333706"
                    /db_xref="SWISS-PROT:P05683"
                    /translation="MPLGVTRKRRDPATKLTNIGNALREONDRLSRAETERRAE
                    GCAVITLDASSIETSPYODRMKGDIDGLTIRBOGOVPIIYRPPSPDGRKYOYAF
                    HRLRAVSELGLPRAVARELTDEOVVAAQGGENNEREDLTETKARFNRHLNRQFSR
                    ELVIAMSIDKSNLSKMLLDVLPSELDAAGAPGVGRPMOOLAEELIEVYSSPAD
                    VAKYAMSEVOALPSAERKAVIASIKPSRVARGPEVATPDGTRIRIAQVOTSKAKLE
                    ITIDRKATPDFAFVLDHYPALYQYAHANQKRKG"
                    complement(1625..1628)
                        /note="put. RNA binding site"
                    2670..2705
                        /note="put. RNA binding site"
                    2710..3927
                        /note="put. rep C (AA 1-403)"
                        /codon_start=1
                        /transl_table=11
                        /protein_id="CAA28532.1"
                        /db_xref="PID:g39119"
                        /db_xref="GI:39119"
                        /db_xref="SWISS-PROT:P05684"
                        /translation="MQGTGSVTPTPGRRPMTALVPRQIALADITOGKTAQKAFRD
                        ASAPAMELIGTQNSLAVIDALSTFHPREELRQALIVFSSNAQLALAHAGAGATLR
                        RHIAIVESGLIVRKDSANGKRYARKDGQIERAFGLSPILARSEELAMAOQVM
                        ABRAAFPMAKESLITCRDVRKRLITLAAMEGEGMOAVEYVELVGIIPAPILAD
                        VSILEEMNLOEIIINRLIENDSNNTNAOSEHIONSKEPSVNELEPSEKED
                        GAKPSRIDARSEPTIKAPRLGWTIKACPTIGVSGSAVSRDIMSAAVYVRSILG
                        SPSAYODACENAGPEMAAAMACILERANFINSPGGLTLDLTRSELGKFSIGPIMA
                        LKASQGTLRFG"
                        3951..3960
```

```

repeat_unit      /note="imp. inverted repeat A"
                  3961..3976
RBS              /note="imp. inverted repeat A'"
                  3980..3986
                  /note="put. rRNA binding site"
                  3992..4255
CDS              /note="put. ORF 2 (AA 1-87)"
                  /codon_start=1
                  /transl_table=11
                  /protein_id="CAA28533.1"
                  /db_xref="PID:39120"
                  /db_xref="GI:39120"
                  /db_xref="SPTREMBL:Q44438"
                  /translation="MSVFMIDRPSDSSTKRALGAILTLNKHGCPMPYVDRESEPA
                  LHAAVFCRRHVAHCLEFALGIVDSVIELADKAITHTEDD"
                  4255..>4638
                  /note="put. ORF 3 (AA 1-128)"
                  /codon_start=1
                  /transl_table=11
                  /protein_id="CAA28534.1"
                  /db_xref="PID:39121"
                  /db_xref="GI:39121"
                  /db_xref="SPTREMBL:Q44439"
                  /translation="MADEIOWPDGTAAEVEPAPRGRAKAPAPETARQSPKSV
                  KPTKGLSNREKLEKIGQIEAOVAGATLKDAVKIVGISVQIYQMKRAAVQPVSONP
                  AVSVVDDELGFIOLEENRRLKRL"
BASE COUNT      1129 a 1191 c 1348 g 970 t
ORIGIN
Query Match      100.0%; Score 10; DB 1; Length 4638;
Best Local Similarity 100.0%; Pred. No. 2.7e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCCACACAG 10
Db 4235 ACCCACACAG 4244

RESULT 7
LOCUS BACCMCASE 1928 bp DNA BCT 02-FEB-1999
DEFINITION B.subtilis carboxymethyl cellulase (CMCase) gene.
ACCESSION D01057
NID 9216387
VERSION D01057.1 GI:216387
WORDS CMCase.
SOURCE Bacillus subtilis (strain:BSE616) DNA.
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
Bacillus.
REFERENCE 1 (bases 1 to 1928)
AUTHORS Park,S.H., Kim,H.K. and Pack,M.Y.
TITLE Characterization and structure of the cellulase gene of Bacillus
subtilis BSE616
JOURNAL Agric. Biol. Chem. 55 (2), 441-448 (1991)
MEDLINE 9129280
COMMENT Submitted (10-JUN-1991) to DDBJ by:
Seung H. Park
Genetic Engineering Research Institute
P.O. Box 17
Taedok Science Town
Taejeon, 305-606
Korea
Phone: 042-829-4312.
Location/Qualifiers
1..1928
/organism="Bacillus subtilis"
/strain="BSE616"
/db_xref="taxon:1423"
-35_signal 140..145
-10_signal 163..166
RBS 209..215

```

```

CDS              228..1727
                  /note="CMCase"
                  /codon_start=1
                  /transl_table=11
                  /protein_id="BAA00859.1"
                  /db_xref="PID:d1001323"
                  /db_xref="PID:9216388"
                  /db_xref="GI:216388"
                  /translation="MKRSISITFICLLITLTMGGMALSPASAGTKTPVAKNQSLI
                  KGTOLVNRDGRAVOLKGISHLQWYGEYVVKDSIKMLRDMDGIIIVFRAMAYTADGSI
                  IDNPSVKMKKEAVEAAKELGIYIIDHILINDGNPNONKEKAEFEKMSLTYGNT
                  NVIEIANEPNGDVNMKREDIPYAEVLSVIRKNDPDIITVGTGTSQDVNDADDO
                  LKDAVMDALHFAVAGTNGOFLRDRANVLSKGAFLPVEMGTSPASGSGVPELOSRE
                  WKIKDSRTISVWMNNSDKOESSALKPGSKTGGFRLSLSASGTFVRENITCTKD
                  STKIPETPADKPTQENGISVQYRAGGDSNMSNOIRPQLOIRKNGNTTVLKDVTAR
                  YWNAKNKGNQNDVDYDOLGCGNVTYKRVTLHKRQGDVTELEGFKNGTLPAGASTG
                  NIOLRHNDWNSNAQSGDYFFKSNFETKTKITLYDQKLIWGTBN"
BASE COUNT      661 a 360 c 452 g 455 t
ORIGIN
repeat_unit      /rpt_type=inverted
                  1757..1764
repeat_unit      /rpt_type=inverted
                  1757..1764
BASE COUNT      661 a 360 c 452 g 455 t
ORIGIN
Query Match      100.0%; Score 10; DB 1; Length 1928;
Best Local Similarity 100.0%; Pred. No. 3.2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCCACACAG 10
Db 1268 ACCCACACAG 1277

RESULT 8
LOCUS BC170DEGR 26170 bp DNA BCT 24-JUN-1998
DEFINITION B.subtilis DNA (26.2 kb fragment; 170 degree region).
ACCESSION Z73234
NID 91405443
VERSION Z73234.1 GI:1405443
KEYWORDS acnltase; a1st gene; ccdA gene; endo-1,4-beta glucanase;
endo-1,4-beta xylanase; gr1A gene; gr1B gene; t1pA gene;
transketolase; yneA gene; yneB gene; yneC gene; yneF gene; yneI
gene; yneJ gene; yneK gene; yneL gene; yneP gene; yneQ gene; yneR
gene; yneS gene; yneT gene; yneU gene; yneV gene; yneW gene; yneX
gene; yneY gene; yneZ gene.
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
Bacillus.
REFERENCE 1 (bases 1 to 26170)
AUTHORS Rose,M. and Entlan,K.D.
TITLE New genes in the 170 degrees region of the Bacillus subtilis genome
encode DNA gyrase subunits, a thiorodoxin, a xylanase and an amino
acid transporter
JOURNAL Microbiology 142 (Pt 11), 3097-3101 (1996)
MEDLINE 97124194
COMMENT Direct Submission
Submitted (19-JUN-1996) Rose M., Johann Wolfgang
Goethe-Universitaet Frankfurt, Institut fuer Mikrobiologie,
Marie-Curie-Str. 9, Geb. N250 Frankfurt/M. GERMANY D-60439
Overlapping sequence: X87845 (bases 3016-6779).
Location/Qualifiers
1..26170
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
82..399
/gene="yneA"
82..399
/gene="yneA"

```

```

/codon_start=1
/transl_table=11
/product="ynef"
/protein_id="CAA97614.1"
/db_xref="PID:e249640"
/db_xref="PID:g1405444"
/db_xref="GI:1405444"
/db_xref="GI:1405444"
/db_xref="SPTREMBL:O45056"
/translation="MIMSKEIIEVGLFTVILISAVILMLSTYSQCEINQYKIEVQO
GDTLMSIADQVADTKRKINKNDIEIWWADKNQLOQSDIQDGLVPLKKKHODAYELA
TVR"
418..1071
/gene="ynef"
418..1071
/gene="ynef"
/feature="similar to gamma-delta resolvase"
/codon_start=1
/transl_table=11
/product="ynef"
/protein_id="CAA97615.1"
/db_xref="PID:e249641"
/db_xref="PID:g1405445"
/db_xref="GI:1405445"
/db_xref="SPTREMBL:O45057"
/translation="MKALITAVSTNKQOENSLKROEBELTALAENGMEVYKYSIE
KASGEMRBDGVFELDELKNDIDVILVQDETIRGKNAKTALHLCYRQGVYTT
AHRGELSEADSVLSEIVSEYQRIHNNKIRGKRAVKNQFQKNLKNHGN
SGKEIEVPSISIVLRANKLTFEAIATLRGQYDSKATVHRFOEYIENEETAE"
1537..3540
/gene="ckta"
1537..3540
/gene="ckta"
/codon_start=1
/transl_table=11
/product="transketolase"
/protein_id="CAA97616.1"
/db_xref="PID:e249642"
/db_xref="PID:g1405446"
/db_xref="GI:1405446"
/db_xref="SWISS-PROT:P45694"
/translation="MDTIEKKSVAITRILSIDAIEKANSQHGPMGAAPNAVITLTK
FMVSPANPQWNRDRFVLSAGHSALISMLHLSGPDLSIEDLKGFQMSKTFGHP
EGRHAGVATGTPGOGIIMAVGALIEBHAEIYNNDSFVVDHITYTICGDDLM
EGISSEASLGHLDGRLIVYDSNDISLDDLDLRSSENVKQREEMNEVLVDE
GNNIEELTAAIEKARONKPKTILIEVKTIGFSPNAGTSGVHGAPELGESEKLTKE
AYATVEEDFVPSSEVYEHFAVAVKSEKKEQENNAQFAKKEVYPLAEQLEIAIK
GELPKDMQEVVYVEKSSILASRASSGEVLNGLAKIKPFVGGASDLAGSNKTTIKNA
GDTFVADYSGKMFRCVREPMAGALNGLHGLRVRFGCTFVPSDLRAIRIALAL
MGLPTVYFTHDSIIVGEDGPTHEPVEQLASIRAPNLSLRPADGNETAAAKLAVQ
SDHPETALVLRONLPTIDQISEALAVEKGAIVYKSKNETPDALLIASGSEVGLA
IEQAELAKENIDVSVMPMDRERKQSDERKNEVLPAQVKKRLAIEMGSSFGWGY
TGLEGDVIGIDRFASAPETIINEYGSVPVVRKALINK"
3693..4139
/gene="ynef"
3693..4139
/gene="ynef"
/feature="identical to yoxf (from acc. no. X87845)"
/codon_start=1
/transl_table=11
/product="ynef"
/protein_id="CAA97617.1"
/db_xref="PID:e249643"
/db_xref="PID:g1405447"
/db_xref="GI:1405447"
/db_xref="SWISS-PROT:P45707"
/translation="MERNYTYTLKEEFANHYFGESYFELFQDIWHTSLKQOYEM
IEKQIOTIOTPIIHMHORLKNMKNIDYQOLDITRIALPKAGHATFMKKEHMI
IVASGDYEAETIFFEVLKRVSPCLAMDPSKRYMLNPKYERNV"
4225..4443
/gene="ynef"
4225..4443
/gene="ynef"
/feature="identical to yoxg (from acc. no. X87845)"

```

```

/codon_start=1
/transl_table=11
/product="ynef"
/protein_id="CAA97618.1"
/db_xref="PID:e249644"
/db_xref="PID:g1405448"
/db_xref="GI:1405448"
/db_xref="GI:1405448"
/db_xref="SWISS-PROT:P45708"
/translation="MTLWVGILVGVALLIGVALGFFIARKYMSYLKKNPINEOML
RMMAMQMKRPSQKINQMAKMNQET"
4910..5617
/gene="ccda"
4910..5617
/gene="ccda"
/feature="cytochrome C-type biogenesis protein"
/codon_start=1
/transl_table=11
/product="ccda"
/protein_id="CAA97594.1"
/db_xref="PID:e249645"
/db_xref="PID:g1405449"
/db_xref="GI:1405449"
/db_xref="SWISS-PROT:P45706"
/translation="MGDVNFTITGAGLSLSPCCLPYPAFLSYTTCVSMQVYKE
KLLOKRSLEFTLGLFVSIFALGYSFISLFRDHDALROIGALLILGFI
TLGVREPMKKERRIHRKHPSCGLSVLGMFAAGWPCPTPIIAAVITLGTNP
GSAPVYMLYVLGAVPFLSLFTKLKWKIRKNQFLMKAGVIMIVIGVLLFFNMN
SLIILSLDIEGSGTGT"
5706..6068
/gene="ynef"
5706..6068
/gene="ynef"
/feature="response regulator"
/feature="identical to yoxh (from acc. no. X87845)"
/codon_start=1
/transl_table=11
/product="ynef"
/protein_id="CAA97595.1"
/db_xref="PID:e249646"
/db_xref="PID:g1405450"
/db_xref="GI:1405450"
/db_xref="SWISS-PROT:P45709"
/translation="MTRVLVVDKFRVAKIRELLEBANTIIAGEADGQADLYRK
LRPLVMDITMPVKNKIKALRDILTFDPRAKYMCTAMQOQRIVTEBAIEGADFTV
KPFETVILEVSRVMGH"
6147..6638
/gene="ynef"
6147..6638
/gene="ynef"
/feature="identical to yoxi (from acc. no. X87845)"
/codon_start=1
/transl_table=11
/product="ynef"
/protein_id="CAA97596.1"
/db_xref="PID:e249647"
/db_xref="PID:g1405451"
/db_xref="GI:1405451"
/db_xref="SWISS-PROT:P45710"
/translation="MISMTIIVSIIIVLAVAVVTRIKSSDKPSPKIIILPPIIM
STGLAMFLFPVFWTGAFLFELFGLVIFSLFKTSKEIKNNEIYMKRSKAFVFI
VGLVIRIVKMSILSTSIDGALSGMFIALFGMIVPWRIAMYLSYKRLHNELOSSNI
QMN"
6669..7097
/gene="ynef"
6669..7097
/gene="ynef"
/feature="identical to yoxj (from acc. no. X87845)"
/codon_start=1
/transl_table=11
/product="ynef"
/protein_id="CAA97597.1"
/db_xref="PID:e249648"
/db_xref="PID:g1405452"

```

```

gene
CDS
    /db_xref="GI:1405452"
    /db_xref="SWISS-PROT:P45711"
    /translation="MLEGFWLFIEMVIMVLSIGFEMERKRLKLPEDGRSE
    LPMODYIENSRLNDEKOFLELTAPEPELFRDAKAKIAGKIGELALKENAKI
    DOGLMKGYIATPKRDHTEFLKRLDKRIDEPYOTLIK"
    complement(7331..7723)
    /gene="yneL"
    /complement(7331..7723)
    /note="similar to heat shock proteins"
    /gene="yneL"
    /note="yneL"
    /codon_start=1
    /transl_table=11
    /product="yneL"
    /protein_id="CAA97598.1"
    /db_xref="PID:g249649"
    /db_xref="PID:g1405453"
    /db_xref="GI:1405453"

Query Match      100.0%; Score 10; DB 1; Length 26170;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAG 10
Db 23341 ACCCACACAG 23350

RESULT 9
LOCUS BSL1GLUC 2314 bp DNA BCT 30-JAN-1995
DEFINITION B.subtilis gene for cellulase.
ACCESSION X67044
VERSION X67044.1 GI:39776
KEYWORDS cellulase; endo-1,4-beta-glucanase.
SOURCE Bacillus subtilis.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
          Bacillus.
REFERENCE 1 (bases 1 to 2314)
AUTHORS Lindahl,V., Aa,K. and Tromsø,M.A.
TITLE Direct Submision
JOURNAL Submitted (22-JUN-1992) V. Lindahl, Agricultural University of
NORWAY, Dept of Biotechnological Sciences, PO Box 40, 1432 AS,
NORWAY
2 (bases 1 to 2314)
Lindahl,V., Aa,K. and Tromsø,M.A.
Nucleotide sequence of an endo-beta-1,4-glucanase gene from
Bacillus subtilis CK-2
JOURNAL Antonie Van Leeuwenhoek 66 (4), 327-332 (1994)
FEATURES
    source
        1..2314
            location/Qualifiers
            /organism="Bacillus subtilis"
            /strain="CK-2"
            /db_xref="taxon:1423"
            /clone_lib="pGEM CEL 101"
            719..724
            742..747
            789..794
            807..2306
            /EC_number="3.2.1.4"
            /note="cellulase"
            /codon_start=1
            /transl_table=11
            /product="endo-beta-1,4-glucanase"
            /protein_id="CAA47429.1"
            /db_xref="GI:39777"
            /db_xref="PID:g39777"
            /db_xref="SWISS-PROT:P10475"
            /translation="MKRSISIFICILITLITMGGMISPSAAGTKTPVAKNQLSI
            KGTOLYNDRKAVOLKGISSHGLQYTGIVKDSIKLRDMDGLTFVFAAYTADGGT
            IDNFSVKNKVAEVAKEELGIYIIDWHIINDGNPNONKEKAEFFREMSLSYGNTP

```

```

BASE COUNT      775 a      437 c      524 g      578 t
ORIGIN
Query Match      100.0%; Score 10; DB 1; Length 2314;
Best Local Similarity 100.0%; Pred. No. 3.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAG 10
Db 1847 ACCCACACAG 1856

RESULT 10
LOCUS BSBGLUC2 2435 bp DNA BCT 30-MAR-1995
DEFINITION Bacillus subtilis endo-beta-1,4-glucanase gene.
ACCESSION X04689
VERSION X04689.1 GI:39823
KEYWORDS beta-glucanase; endo-beta-1,4-glucanase; glucanase; unidentified
          reading frame.
SOURCE Bacillus subtilis.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
          Bacillus.
REFERENCE 1 (bases 1 to 2435)
AUTHORS Mackay,R.M., Lo,A., Willick,G., Zuker,M., Baird,S., Dove,M.,
TITLE Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene
JOURNAL Nucleic Acids Res. 14 (22), 9159-9170 (1986)
FEATURES
    source
        1..2435
            location/Qualifiers
            /organism="Bacillus subtilis"
            /strain="PAP115"
            /db_xref="taxon:1423"
            <1..428
            /note="unidentified reading frame (142Aa) (1 is 2nd base
            in codon)"
            /codon_start=3
            /transl_table=11
            /protein_id="CAA28391.1"
            /db_xref="PID:g757741"
            /db_xref="GI:757741"
            /db_xref="SPTRMBL:045525"
            /translation="IOVTOALNHHHIGNAPFIFVAMFLFAFSSVGVNYGETNIE
            FIKTSKTLNIIYRIAVIYAVVIGSLSGFOIWMADDFMGLMILNIIIVIALSNVAY
            KVVYKDYAKORKOGIDPYFKAKNIPGLKNAITWDEKGA"
            443..473
            /note="put. transcription terminator"
            443..456
            /note="inverted repeat A"
            460..473
            /note="inverted repeat A"
            643..648
            /note="put. -35 region"
            666..671
            /note="put. -10 region"
            712..726
            /note="put. rRNA binding site"
            731..817
            /product="signal peptide (AA -29 to -1)"
            731..2230
            /note="precursor polypeptide (AA -29 to 470)"
            /codon_start=1
            /transl_table=11

```

```

/protein_id="CAA28392.1"
/db_xref="PID:g39824"
/db_xref="GI:39824"
/db_xref="SWISS-Prot:P10475"
/translation="MKRSISIFICLLITLTLTLMGMIASPAAGTKTPVAKNGOLSI
KGTOLVNRDGAVALDKGISSHGLOWGEYVKNDSIKMLRDMGTFVPAAMYADGGY
IDNPSVKNKVEAVEAKELGIYIIDMHILNDGPNONKREKAEFEKMSILGNTP
NVIETIANEPNGDVNMRKDIKPAEVEISVIRKNDPNIIIVGTGTSQDVNDADDO
LKDNVMTALHFAAGTHGQFLRDKANALSKGAPFVTEWGTSDASGNGVFLDQSR
WLKYLDSKTSISVNMNLSDKQESSALKPGASKGWRMLSDLSAGTFVENIIGTRD
STKDIPTSPKDKPTQENGISVQYRAGDSMNSQIROPQIKNNGNTVYDLKVTAR
YWKAKKNGONFDCDYAOIGGANTHGFVTLHKPKOGADTYLELGFKNGTLPAGASTG
NIOLRLHNDMSVNAQSGDSYFESKNTFKTKITLIDGKLLINGTEPN"
mat_peptide
818..2227
/product="endo-beta-1,4-glucanase (AA 1-470)"
repeat_unit
2247..2254
/translation="MKRSISIFICLLITLTLTLMGMIASPAAGTKTPVAKNGOLSI
KGTOLVNRDGAVALDKGISSHGLOWGEYVKNDSIKMLRDMGTFVPAAMYADGGY
IDNPSVKNKVEAVEAKELGIYIIDMHILNDGPNONKREKAEFEKMSILGNTP
NVIETIANEPNGDVNMRKDIKPAEVEISVIRKNDPNIIIVGTGTSQDVNDADDO
LKDNVMTALHFAAGTHGQFLRDKANALSKGAPFVTEWGTSDASGNGVFLDQSR
WLKYLDSKTSISVNMNLSDKQESSALKPGASKGWRMLSDLSAGTFVENIIGTRD
STKDIPTSPKDKPTQENGISVQYRAGDSMNSQIROPQIKNNGNTVYDLKVTAR
YWKAKKNGONFDCDYAOIGGANTHGFVTLHKPKOGADTYLELGFKNGTLPAGASTG
NIOLRLHNDMSVNAQSGDSYFESKNTFKTKITLIDGKLLINGTEPN"
terminator
2247..2268
/translation="MKRSISIFICLLITLTLTLMGMIASPAAGTKTPVAKNGOLSI
KGTOLVNRDGAVALDKGISSHGLOWGEYVKNDSIKMLRDMGTFVPAAMYADGGY
IDNPSVKNKVEAVEAKELGIYIIDMHILNDGPNONKREKAEFEKMSILGNTP
NVIETIANEPNGDVNMRKDIKPAEVEISVIRKNDPNIIIVGTGTSQDVNDADDO
LKDNVMTALHFAAGTHGQFLRDKANALSKGAPFVTEWGTSDASGNGVFLDQSR
WLKYLDSKTSISVNMNLSDKQESSALKPGASKGWRMLSDLSAGTFVENIIGTRD
STKDIPTSPKDKPTQENGISVQYRAGDSMNSQIROPQIKNNGNTVYDLKVTAR
YWKAKKNGONFDCDYAOIGGANTHGFVTLHKPKOGADTYLELGFKNGTLPAGASTG
NIOLRLHNDMSVNAQSGDSYFESKNTFKTKITLIDGKLLINGTEPN"
repeat_unit
2261..2268
/translation="MKRSISIFICLLITLTLTLMGMIASPAAGTKTPVAKNGOLSI
KGTOLVNRDGAVALDKGISSHGLOWGEYVKNDSIKMLRDMGTFVPAAMYADGGY
IDNPSVKNKVEAVEAKELGIYIIDMHILNDGPNONKREKAEFEKMSILGNTP
NVIETIANEPNGDVNMRKDIKPAEVEISVIRKNDPNIIIVGTGTSQDVNDADDO
LKDNVMTALHFAAGTHGQFLRDKANALSKGAPFVTEWGTSDASGNGVFLDQSR
WLKYLDSKTSISVNMNLSDKQESSALKPGASKGWRMLSDLSAGTFVENIIGTRD
STKDIPTSPKDKPTQENGISVQYRAGDSMNSQIROPQIKNNGNTVYDLKVTAR
YWKAKKNGONFDCDYAOIGGANTHGFVTLHKPKOGADTYLELGFKNGTLPAGASTG
NIOLRLHNDMSVNAQSGDSYFESKNTFKTKITLIDGKLLINGTEPN"
BASE COUNT      819 a      450 c      556 g      610 t
ORIGIN
Query Match      100.0%; Score 10; DB 1; Length 2435;
Best Local Similarity 100.0%; Pred. No. 3.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 ACCGACAG 10
Db      1771 ACCGACAG 1780

```

```

/standard_name="1,4-beta-D-glucan-4-glucanohydrolase"
/EC_number="3.2.1.4"
/function="hydrolysing carboxymethyl cellulose"
/note="Sequence has been found completely identical with
that published by R.M.McKay et al. (1986) NAR 14,
9159-9170 from B subtilis Pap15 (nt452-nt2251)"
/codon_start=1
/label="table-11"
/evidence="experimental"
/product="endo-1,4-beta-glucanase"
/protein_id="CAA82317.1"
/db_xref="PID:g509267"
/db_xref="GI:509267"
/db_xref="SWISS-Prot:P10475"
/translation="MKRSISIFICLLITLTLTLMGMIASPAAGTKTPVAKNGOLSI
KGTOLVNRDGAVALDKGISSHGLOWGEYVKNDSIKMLRDMGTFVPAAMYADGGY
IDNPSVKNKVEAVEAKELGIYIIDMHILNDGPNONKREKAEFEKMSILGNTP
NVIETIANEPNGDVNMRKDIKPAEVEISVIRKNDPNIIIVGTGTSQDVNDADDO
LKDNVMTALHFAAGTHGQFLRDKANALSKGAPFVTEWGTSDASGNGVFLDQSR
WLKYLDSKTSISVNMNLSDKQESSALKPGASKGWRMLSDLSAGTFVENIIGTRD
STKDIPTSPKDKPTQENGISVQYRAGDSMNSQIROPQIKNNGNTVYDLKVTAR
YWKAKKNGONFDCDYAOIGGANTHGFVTLHKPKOGADTYLELGFKNGTLPAGASTG
NIOLRLHNDMSVNAQSGDSYFESKNTFKTKITLIDGKLLINGTEPN"
BASE COUNT      609 a      336 c      416 g      439 t
ORIGIN
Query Match      100.0%; Score 10; DB 1; Length 1800;
Best Local Similarity 100.0%; Pred. No. 3.3e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 ACCGACAG 10
Db      1320 ACCGACAG 1329

```

```

RESULT 11
BSEQLSG      1800 bp      DNA      BCT      26-MAR-1998
LOCUS      B:subtilis (168 trpC2) egIs gene for endo-1,4-beta-glucanase.
DEFINITION
ACCESSION      229076
NID      9509266
VERSION      229076.1 GI:509266
KEYWORDS      egIs gene; endo-1,4-beta-glucanase.
SOURCE      Bacillus subtilis.
ORGANISM      Bacillus subtilis; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
REFERENCE
1 (bases 1 to 1800)
Mackay,R.M., Lo,A., Willick,G., Zuker,M., Baird,S., Dove,M.,
Moranell,F. and Seligy,V.
Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene
Nucleic Acids Res. 14 (22), 9159-9170 (1986)
2 (bases 1 to 1800)
Borriass,R.
Direct Submission
Submitted (14-DEC-1993) Borriass R., Humboldt Univ. zu Berlin,
Biologie, Warschauerstrasse 43, Berlin, Deutschland, D-10243
3 (bases 1 to 1800)
Wolf,M., Geczi,A., Simon,O. and Borriass,R.
Genes encoding xylan and beta-glucan hydrolysing enzymes in
Bacillus subtilis: characterization, mapping and construction of
Microbiolology 141 (Pt 2), 281-290 (1995)
95219081
JOURNAL
MEDLINE
FEATURES
SOURCE
Location/Qualifiers
1..1800
/organism="Bacillus subtilis"
/strain="168 trpC2"
/db_xref="taxon:1423"
280..1779
/gene="egIs"
280..1779
/gene="egIs"
CDS

```

```

RESULT 12
BSUB0010      233780 bp      DNA      BCT      26-NOV-1997
LOCUS      Bacillus subtilis complete genome (section 10 of 21): from 1781201
DEFINITION
ACCESSION      299113 AL009126
NID      92634090
VERSION      299113.1 GI:2634090
KEYWORDS
SOURCE      Bacillus subtilis.
ORGANISM      Bacillus subtilis; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
REFERENCE
1 (bases 1 to 233780)
Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
Borriass,R., Bounisier,L., Brans,A., Braun,M., Bridnell,S.C.,
Bron,S., Brouillet,S., Brusch,C.V., Caldwell,B., Capuano,V.,
Carter,N.M., Choi,S.K., Codani,J.J., Conerton,I.F., Cummings,N.J.,
Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
Emmerson,P.T., Entlan,K.D., Errington,J., Fabeet,C., Ferrara,E.,
Folger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
Gallerson,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
Grandi,G., Guisepi,G., Guy,B.J., Haga,K., Halech,J., Harwood,C.R.,
Haupt,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
Klaer-Bianchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
Masuda,S., Mauel,C., Medigue,C., Medina,N., Melado,R.P.,
Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
Ogawa,K., Ogilwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
Portetelle,D., Porwollik,S., Prescott,A.M., Prescan,E., Pujic,P.,
Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
Rivolta,C., Rochat,E., Roche,B., Rose,M., Sadate,Y., Sato,T.,
Scanlan,E., Schleich,S., Schroeter,R., Scottone,F., Sekiguchi,J.,

```





255 .9292  
/gene="pksg"  
/function="Involved in polypeptide synthesis"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAB13597.1"  
/db\_xref="pid:el183372"  
/db\_xref="pid:g2634097"  
/db\_xref="GI:2634097"  
/db\_xref="SWISS-PROT:P40830"  
/translation="MLITCSESGIDFGKSLSTYIHEYLGILNRNCRLEPVKQACYSGTAR  
GFOCAVNIILSQTSPGAKALVIASDIISFLLAEGDALSEMSYAEPGAGAAVALV  
GENEVEFIDPGANGYGYEVMDCRPIPDSEAGDSIDDCOFTLEOKRVR  
PGAVQDFQFOYLAYHTPGVAKGKAPRMKAKYKVSIGETFDLFRVKGVLNACCR  
VGNIMGAAFLTALASTIDQGFDPFKRGICSYSGCCSESYSGITTPQGGERQRTFR  
IEKHLDRYQLSMENEYELFFKSGVNRFGTRNVKLDPEMIPGIMQSTQEKPRFLLEET  
SEFRKRWYIS"  
SEFRRKRWYIS"  
9280 .10059  
/gene="pksh"  
9280 .10059

Query Match 100.0%; Score 10; DB 1; Length 233780;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ACCCACACAG 10  
|||||  
Db 159720 ACCCACACAG 159729

RESULT 13  
BSUB0017  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

BSUB0017 217420 bp DNA BCT 26-NOV-1997  
Bacillus subtilis complete genome (section 17 of 21): from 3197001  
to 3414420.  
299120 AL009126  
92635613  
299120.1 GI:2635613

Bacillus subtilis  
Bacillus subtilis  
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;  
Bacillus.  
1 (bases 1 to 217420)  
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,  
Azevedo, V., Bartorelli, M.G., Bessières, P., Bolotin, A., Borcher, S.,  
Borries, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C.,  
Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V.,  
Carter, N.M., Choi, S.K., Codan, J.J., Connetton, I.F., Cummings, N.J.,  
Daniel, R.A., Denzot, F., Devigne, R.M., Dusterhoft, A., Ehrlich, S.D.,  
Emmerson, P.T., Entlian, R.D., Errington, J., Fabret, C., Ferrari, E.,  
Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Gallizzi, A.,  
Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golightly, E.J.,  
Grandi, G., Giuseppe, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R.,  
Hennaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,  
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,  
Klaier-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,  
Koenigstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,  
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,  
Masuda, S., Meuel, C., Medigue, C., Medina, N., Mellado, R.P.,  
Minudo, M., Moenst, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,  
Ogawa, K., Ogata, A., Oudega, B., Park, S.H., Paro, V., Pohl, T.M.,  
Portetalle, D., Porwollik, S., Prescott, A.M., Prescan, E., Pujic, P.,  
Punelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,  
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadale, Y., Sato, T.,  
Scanlan, E., Schleich, S., Schreier, R., Scoffone, F., Sekiguchi, J.,  
Sekowska, A., Seror, S.J., Serri, P., Shin, B.S., Soldo, B.,  
Sorochni, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,  
Takeuchi, A., Yamakoshi, A., Tanaka, T., Terpestra, P., Tognoni, A.,  
Tosato, V., Uchikawa, S., Vandenhof, M., Vannier, F., Vassart, A.,  
Viat, A., Wambut, R., Wedler, E., Wedler, H., Weltzenegger, T.,  
Winters, P., Wipar, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,  
Yoshida, K., Yoshikawa, H., F., Zumststein, E., Yoshikawa, H. and

FEATURES	Source
TITLE	Danchin, A.
JOURNAL	The complete genome sequence of the gram-positive bacterium
MEDLINE	<i>Bacillus subtilis</i>
REFERENCE	Nature 390 (6657), 249-256 (1997)
AUTHORS	98044033
TITLE	2 (bases 1 to 217420)
JOURNAL	Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Direct Submission	Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr. Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48
Location/Qualifiers	1. 217420
/organism="Bacillus subtilis"	
/strain="168"	
/db_xref="taxon:1423"	
complement(3. 1277)	
/gene="yulB"	
complement(3. 1277)	
/gene="yulE"	
/function="unknown"	
/note="similar to L-rhamnose isomerase"	
/codon_start=1	
/transl_table=1	
/protein_id="CAB15107.1"	
/db_xref="PID:e1184196"	
/db_xref="PID:g2635614"	
/db_xref="GI:2635614"	
/db_xref="SPTRMBL:005264"	
/translation="MIRKASIDSAKQAEYERKMGIDVEBALROLQEQVPSIHQMOGDIE GFENKELSGGIDVGNYPYKAQTPPELRDLKALSLIPGKHVNLHAIETNREE AVEDEKELPQHFENWYKAKNLGIDFNPTLPSEKAAGLITSHDPDIPREPIRKH CICRRIGEVKPEKELGCPYCLNIMIPGYKDIPDRLPRLKRESIDRLPSEISRO HNDSITSKLPGLCSBESYVYGSEFTYLAFLYLNHKLCLLDTGHHPEPTYKSSML LYDKLALHVSRPYRWDSDHVVVLDLREIATLEIVRNHALERVAIGLDFFDASINVL AAWITGRNNIKALYVLLPGLYKLOJEGRYTERLALMEEFKTYPGAIDSDYCE OMGYPVKAMLYDLKEVEQVLLKRAKSSPIV"	
complement(1303. 1617)	
/gene="yulD"	
complement(1303. 1617)	
/gene="yulC"	
/function="unknown"	
/note="similar to hypothetical proteins"	
/codon_start=1	
/transl_table=1	
/protein_id="CAB15108.1"	
/db_xref="PID:e1184197"	
/db_xref="PID:g2635615"	
/db_xref="GI:2635615"	
/db_xref="SPTRMBL:005263"	
/translation="MKRKASIMYVHQDKYEYIKORHDDIPENAEAKAKGAHHYSIF LDETSGLRFLVLEIDEDKMKKADTVEVCGRWKMSKAPLMKNTSDFSVAIDLKEVY LD"	
complement(1635. 3092)	
/gene="yulC"	
complement(1635. 3092)	
/gene="yulC"	
/function="unknown"	
/note="similar to rhannulokinase"	
/codon_start=1	
/transl_table=1	
/protein_id="CAB15109.1"	
/db_xref="PID:e1184198"	
/db_xref="PID:g2635616"	
/db_xref="GI:2635616"	
/db_xref="SPTRMBL:005262"	
/translation="MITYTADVGSAGSRIMVGELENEGLDIOEIHPRANGESORDGHC LMIDHLLKQILOGLOKVRVIGEHCVGIDTNAVVDVLLDGGDRLEAISRDRT DHTIDKLEHTLSKRAIITOKTGIOFOPNTIYQLEEDRELKTKDKTMMAPDLGICL TGRAVEITNVSTQLVNVTGNDPELDAVSLVEQFAPLTPGCGELKLRNEWP DYDLPAKCVTVAITHDASAVIAPGVNDGVAISSGTWSLIGVENTPPIITDLALEN	

NVTNRGANNTIRFKNTIGMVIQVQKQLOADYSPQOLAEEAKKTEPFOOTLND  
 KRFLENPMKIEIOHYCROTROKIPRTAGELACCISYLAITAIKELTEITKPI  
 EOHFIIGGARDFLNDLTADSGKAVAGTLEATATISLIMOTIAKEVADIKEARO  
 VVRNFPKIVFTPKIDIRSTIIQSFQOTVALASK"  
 complement(3097. .3873)  
 gene="yulb"  
 complement(3097. .3873)  
 CDS  
 /gene="yulb"  
 /function="unknown"  
 /note="similar to transcriptional regulator (Deor family)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAB15110.1"  
 /db\_xref="PID:e1184199"  
 /db\_xref="PID:g2635617"  
 /db\_xref="GI:2635617"  
 /db\_xref="SWISS-PROT:O05261"  
 /translation="MLVAERQOKIYEIYNMSSIRSELSDFVTEETIRDLKLE  
 KEHKSRSRGAVSIOOKESEHSEFSEITVIEKKATAHAARVAGSDRIIDAST  
 TAWYAKTLPDIETLVITNSKAATEISNKNISVISGGLKESLSPADLARS  
 ETIYVAKTFLSCGFDVNGMSDSEMOALKKRIERSDQITLMAOSSKGNFEFSH  
 IASLDVSRLLITDSGLDPASVKALEDKRVAVPLSKRG"  
 complement(3930. .5999)  
 gene="yuxg"  
 complement(3930. .5999)  
 CDS  
 /gene="yuxg"  
 /function="unknown"  
 /note="alternate gene name: yulA; similar to  
 sorbitol-6-phosphate 2-dehydrogenase"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAB15111.1"  
 /db\_xref="PID:e1184200"  
 /db\_xref="PID:g2635618"  
 /db\_xref="GI:2635618"  
 /db\_xref="SWISS-PROT:P40747"  
 /translation="MKHIMDSERMAOLPKVEELVYRNSLIGSDRTYCMNGGNTSM  
 KTEDEFGRELEVMMVKGSSDLATMAKNAHPSGLDDIPRLKROMPEBEVDTL  
 SHKIDSKHPRISITILHAFLEPHVDHHPDALTISCDNCKQAEIDYGRFV  
 VPYVPGFTLSKMAIEGVANNPHELVMEHGLVMTGETSETCYOKTISIODEY  
 INDRINQHEVFGGRQYPLPEDKROILAGIMPVRCGAVSEKKMILSYDDHDLVE  
 VNSVOAPALSOIGAACPDHLVHTKRPVLYIDMNETODVHLADLISGVDFSE  
 AVFTNODGDOIFESAPRVLLIGIGVNTGMSKAVSGALYRRAIAMKATL  
 CQFVSLHNESEHYEYPLLEIKULAPPEAFSRKVALITGGAGGIGSACRPALE  
 GGHVADLVNIEGAOKIAGEINDAVGRAAVAVKVDKEEDVSAERAAALAGGD  
 IVYVNAAGLATSPEDETSLEKMNAMNVLGCTPLVAREAKOMKHORGSNMFVS  
 KNSYVAGKNASAVSKALETHLARCIAEGEGEIRVNSVLPDAVLOGSALMGSSWR  
 EERAAVYGIPEPQLEHYRKRRTALLVNIYPPEDIALMALFFASSRAEKTCGMIIVDGS  
 VPAAFTR"  
 complement(6109. .8125)  
 gene="t1lpb"  
 complement(6109. .6124)  
 terminator  
 /gene="t1lpb"  
 complement(6137. .8125)  
 CDS  
 /gene="t1lpb"  
 /codon\_start=1  
 /transl\_table=11  
 /product="methyl-accepting chemotaxis protein"  
 /protein\_id="CAB15112.1"  
 /db\_xref="PID:e1184201"  
 /db\_xref="PID:g2635619"  
 /db\_xref="GI:2635619"  
 /db\_xref="SWISS-PROT:P39217"  
 /translation="MGRFIWKQIPKPSIKPLIAFLAVLIPGVLAIFYSQAMNL  
 DRELISAKGVEELNSTLQNKLEKDAIDYSEVYDKDILLKNTLLKEKQY  
 TLNDVGAIVYASDEBKLYKYPDSGPKPGDGMKQAVAEQGVAFEPPTDA  
 TGDIVYISKOAKDSGVIALDNLDEVLTASRKIKGEGFATITGNKVIAPHPI  
 KPGTSGDMTNOVSKREGSEFEYFECKEAKFTNKLKAGIYPISELODAS  
 SPVANTATILCVSIVIGTILITRITPRKLKYSKISSGDIETVTHSKN  
 EFGQLGESFNESASLRVIGVLOTSEVAVSSSELTASAQTSKATEHTTLAIEP  
 SDGNEAOKERLETSSNLSQNEGISKVAQSSITTKSSIQSSSEAGSGELVHEATG  
 QMTKIDISOVQAEVAVKGLTQSDITSILVINGIADQNTLLALNAIEVARAGEYG  
 RGSVVAEVEVRKLAQVADSASKEIGLIEIVREISTLSMFSQVNHVEKGLITDQ

TASFQIYEMTIOISGELONLNATVQESAGSOEVSAYEDISAVAKESSAGIODIA  
 ASAEQDASHEEISSAETIANNAAEELQITKFKIES"  
 complement(8204. .8228)  
 gene="mcpA"  
 complement(8204. .10224)  
 /gene="mcpA"  
 complement(8239. .10224)  
 CDS  
 /gene="mcpA"  
 /function="chemotaxis deficiency toward glucose and  
 alpha-methyl-glucoside"  
 /codon\_start=1  
 /transl\_table=11  
 /product="methyl-accepting chemotaxis protein"  
 Query Match 100.0%; Score 10; DB 1; Length 217420;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 ACCCACACAG 10  
 Db 150765 ACCCACACAG 150774  
 RESULT 14  
 BTU04366  
 LOCUS  
 DEFINITION Bacillus thuringiensis japonensis Bulbul CryIII delta-endotoxin  
 gene, complete cds.  
 ACCESSION U04366  
 NID 9532523  
 VERSION U04366.1 GI:532523  
 KEYWORDS  
 SOURCE Bacillus thuringiensis.  
 ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group;  
 Bacillus/staphylococcus group; Bacillus.  
 REFERENCE 1 (bases 1 to 3483)  
 AUTHORS Sato,R., Takeuchi,K., Ogihara,K., Minami,M., Kaji,Y., Suzuki,N.,  
 Hori,H., Asano,S., Ohba,M. and Iwahana,H.  
 TITLE Cloning, heterologous expression, and localization of a novel  
 crystal protein gene from Bacillus thuringiensis serovar japonensis  
 strain bulbul toxic to scarabaeid insects  
 JOURNAL Curr. Microbiol. 28 (1), 15-19 (1994)  
 MEDLINE 94100786  
 REFERENCE 2 (bases 1 to 3483)  
 AUTHORS Fellelson,J.S.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-DEC-1993) Jerald S. Fellelson, Molecular Biology,  
 Mycogen Corporation, 4980 Carroli Canyon Road, San Diego, CA 92121,  
 USA  
 COMMENT On Aug 27, 1994 this sequence version replaced gi:436836.  
 FEATURES  
 source  
 1..3483  
 /organism="Bacillus thuringiensis"  
 /strain="bulbul"  
 /sub\_species="japonensis"  
 /db\_xref="taxon:1428"  
 /clone\_lib="Library of K. Ogihara"  
 /note="serotype 23"  
 1..3483  
 /codon\_start=1  
 /transl\_table=11  
 /product="CryIII delta-endotoxin"  
 /protein\_id="AA21119.1"  
 /db\_xref="PID:g532524"  
 /db\_xref="GI:532524"  
 /translation="MSPNNQNEVEITDAISPTSVSNSTIRYPIANDQNTLQWAKYD  
 YLKATSETNALERNRQETPSADAVGTGIDITSTISGIGIVLGEVSTLSLIGL  
 LPSNNENWQITFMNREELIDKILDSVSRALADLANRSIAVEYQNLAEWRNP  
 HSTRSALVNERGNAEALIRTNMGSSQQNYETPLPIYQAQASHLVNRVOYIG  
 KEMGYQNDLDLEKQEVSTAYASDHCQVWYAGLNLKGTGAKQWYDNRPREKN  
 VMVLDVLAEPNDARYPLETNAELTRELFTDPGVSYVAGOSTLSLIDYDMIPALP  
 SFTLENLAKPPFTLLDRIHNTSRQNGTIEYINYNMGQALITISYIGSSFNKS



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:33:33 ; Search time 425.19 Seconds  
(without alignments)  
5.884 Million cell updates/sec

Title: US-08-956-518a-89  
Perfect score: 10  
Sequence: 1 ACCCACACAG 10

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

abase : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	100.0	2181	1 003577	Fragment of sequen
2	10	100.0	1109	1 003852	Sequence encoding
3	10	100.0	2181	1 003668	Fragment encoding
4	10	100.0	384	1 N90572	Caprine repeat ele
5	10	100.0	448	1 N90573	Caprine repeat ele
6	10	100.0	2589	1 N90574	Caprine repeat ele
7	10	100.0	618	1 N90766	Sequence of the am
8	10	100.0	517	1 N91654	CDNA sequence of t
9	10	100.0	547	1 N91655	Synthetic tumour n
10	10	100.0	3086	1 N91815	Egr-1-contg. OC3.1
11	10	100.0	1310	1 N91469	Human lymphotoxin
12	10	100.0	453	1 N93191	Human lymphotoxin
13	10	100.0	925	1 N93196	Monkey recombinant
14	10	100.0	618	1 003289	New synthetic huma
15	10	100.0	3344	1 N81442	Clone encoding hum
16	10	100.0	10475	1 N80616	Genomic clone enco
17	10	100.0	618	1 N82251	Variant lymphotoxi
18	10	100.0	1077	1 N91628	Sequence of gamma-
19	10	100.0	1320	1 N80599	Encodes lymphotoxi
20	10	100.0	618	1 006155	CDNA encoding Lym
21	10	100.0	2654	1 N70597	Sequence encoding
22	10	100.0	5522	1 N70062	Sau3A-EcoRI 5.4 kb
23	10	100.0	10475	1 N80643	Sequence of human
24	10	100.0	456	1 N70441	Sequence encoding
25	10	100.0	1310	1 N70439	Sequence encoding
26	10	100.0	492	1 N70440	Sequence encoding
27	10	100.0	447	1 N70111	Sequence encoding
28	10	100.0	951	1 N70255	Sequence of interf
29	10	100.0	1014	1 N70256	Sequence of interf
30	10	100.0	1966	1 N70316	Sequence encoding
31	10	100.0	1530	1 N70318	Sequence encoding
32	10	100.0	907	1 N70905	Human Kallikrein-1
33	10	100.0	738	1 N71049	Sequence encoding
34	10	100.0	2160	1 N91624	5' flanking sequen
35	10	100.0	682	1 Q13586	Encodes amino acid
36	10	100.0	5534	1 Q13574	Ecr gene, DNA enco
37	10	100.0	1863	1 Q14271	D-3 dopaminergic r
38	10	100.0	1863	1 Q14272	D-3 dopaminergic r
39	10	100.0	1337	1 N50046	Sequence encoding
40	10	100.0	15155	1 N50107	DNA encoding Facto
41	10	100.0	1109	1 Q20612	Murine macrophage
42	10	100.0	1109	1 Q20728	Murine macrophage
43	10	100.0	17041	1 Q21065	Genomic DNA of hum

44 10 100.0 626 1 022228  
c 45 10 100.0 6614 1 X26905

## ALIGNMENTS

RESULT 1  
ID 003577 standard; DNA; 2181 BP.  
AC 003577;  
DT 02-AUG-1990 (first entry)  
DE Fragment of sequence encoding Heymann nephritis antigen, gp330.  
KW Heymann nephritis antigen; HNA; gp330; familial  
KM hypercholesterolaemia; membranous glomerulonephritis; ds.  
OS Rattus sp.  
PN EP-358977-A.  
PD 21-MAR-1990.  
PF 19-AUG-1989; 115331.  
PR 23-AUG-1988; US-235211.  
PR 22-FEB-1989; US-313682.  
PA (GEO-) Gen Hospital Corp.  
PI Smith JA, Raychowdhury R, Niles JL;  
DR WPI: 90-084851/12.  
DR P-PSDB: R05533.  
PT Cloned Heymann nephritis antigen, gp330, gene -  
PT used for diagnosis and in treatment of membranous  
PT glomerulonephritis, wounds and gastric ulcers.  
PS Disclosure; P; English.  
CC The gene can be used to transform a host cell providing functional  
CC derivatives, free from impurities.  
SQ Sequence 2181 BP; 533 A; 541 C; 561 G; 526 T;

Query Match 100.0%; Score 10; DB 1; Length 2181;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAG 10  
Db 444 ACCCACACAG 453

RESULT 2  
ID 003852 standard; DNA; 1109 BP.  
AC 003852;  
DT 23-AUG-1990 (first entry)  
DE Sequence encoding macrophage derived inflammatory cytokine, MIP-2.  
KW MIP-2; macrophage-derived inflammatory cytokine; heparin; ds.  
OS Mus musculus.  
FH Key Location/Qualifiers  
FT cds 47..346  
FT mat\_peptide 125..346  
FT /\*tag- a  
FT /\*tag- b  
PN WO9002762-A.  
PD 22-MAR-1990.  
PF 1-SEP-1989; 003798.  
PR 2-SEP-1988; US-240078.  
PR 1-SEP-1989; US-399971.  
PA (CHIR-) Chiron Corp., (UNRO-) Rockefeller Univ.  
PI Wolpe SD, Cerami A, Sherry B, Olson-Tekamp PA;  
DR WPI: 90-115959/15.  
DR P-PSDB: R05790.  
PT Macrophage-derived inflammatory cytokine MIP-2 -  
PT used for developing screening assays, for prodn. of antibodies  
PT and in diagnosis and therapy.  
PS Claim 5; Fig 7; 78pp; English.  
CC MIP-2 may be used to detect idiopathic or invasive stimuli.  
CC Antibodies raised to the cytokine may be used to measure MIP-2  
CC activity and inhibit the production of MIP-2 used in treatment of  
CC inflammation or fever.

Sequence 1109 BP; 320 A; 228 C; 241 G; 320 T;

Query Match  
Best Local Similarity 100.0%; Score 10; DB 1; Length 1109;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAG 10  
DB 735 ACCCACACAG 744

RESULT 3

ID 003668 standard; DNA; 2181 BP.  
AC 003668;  
DT 07-AUG-1990 (first entry)  
DE Fragment encoding gp330, Heymann's nephritis antigen.  
KW Heymann's nephritis antigen; gp330; LDL receptor;  
RN Familial cholesterololemia; FXIIa; ss.  
PN Rattus rattus.  
PD WO9002207-A.  
PF 08-MAR-1990.  
PR 23-AUG-1989; U03621.  
PR 23-AUG-1988; US-235211.  
PR 22-FEB-1989; US-313682.  
PR 22-AUG-1989; US-396697.  
PA (GEHO-) General Hospital Corp.  
PI Smith JA, Raychowdhury R, Miles JL;  
DR WPI: 90-099425/13.  
P-PSDB; R05580.  
PT Recombinant DNA encoding and expressing Heymann's nephritis antigen -  
its use for therapeutic and/or preventative administration.  
PS Claim 3; Fig 8; 65pp; English.  
CC The sequence was derived from two overlapping clones of 1.4 kb and a 2 kb  
CC resp. It encodes three cysteine-rich regions which are homologous to the  
CC LDL receptor binding domain.  
CC Sequence 2181 BP; 533 A; 542 C; 580 G; 526 T;

Query Match  
Best Local Similarity 100.0%; Score 10; DB 1; Length 2181;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAG 10  
DB 444 ACCCACACAG 453

RESULT 4

ID N90572 standard; DNA; 384 BP.  
AC N90572;  
DT 12-MAR-1990 (first entry)  
DE Caprine repeat element GRV.1a(a).  
KW GRV.1a(a); Y-chromosome; ruminants; repeat element; sex determination;  
OS Capra hircus.  
PN WO8907154-A.  
PD 10-AUG-1989.  
PR 27-JAN-1989; AU00029.  
PR 29-JAN-1988; AU-006476.  
PA (ADRI-) Advanced Riverina Holdings Ltd.  
PI Reed KC; Lord EA; Matthea KI; Mann DA; Beaton S; Herr CM; Matthews ME;  
DR WPI: 89-249021/34.  
PT New nucleic acid isolates hybridising only ruminant Y chromosomal DNA -  
useful in hybridisation and polymerase chain reaction tests for sex  
PT determination.  
PS Fig. 8; 20 pp; English.  
CC The sequence is isolated from male goats and hybridises only to the DNA  
CC of cattle, sheep, goats and other ruminants which contain Y-chromosomal  
CC DNA. It can be used for sex determination.  
CC Sequence 384 BP; 113 A; 91 C; 87 G; 93 T;

Query Match  
Best Local Similarity 100.0%; Score 10; DB 1; Length 384;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAG 10  
DB 155 ACCCACACAG 164

RESULT 5

ID N90573 standard; DNA; 448 BP.  
AC N90573;  
DT 12-MAR-1990 (first entry)  
DE Caprine repeat element GRV.1b(a).  
KW GRV.1b(a); Y-chromosome; ruminants; repeat element; sex determination;  
OS Capra hircus.  
PN WO8907154-A.  
PD 10-AUG-1989.  
PR 27-JAN-1989; AU00029.  
PR 29-JAN-1988; AU-006476.  
PA (ADRI-) Advanced Riverina Holdings Ltd.  
PI Reed KC; Lord EA; Matthea KI; Mann DA; Beaton S; Herr CM; Matthews ME;  
DR WPI: 89-249021/34.  
PT New nucleic acid isolates hybridising only ruminant Y chromosomal DNA -  
useful in hybridisation and polymerase chain reaction tests for sex  
PT determination.  
PS Fig. 9; 20 pp; English.  
CC The sequence is isolated from male goats and hybridises only to the DNA  
CC of cattle, sheep, goats and other ruminants which contain Y-chromosomal  
CC DNA. It can be used for sex determination.  
CC Sequence 448 BP; 132 A; 104 C; 103 G; 109 T;

Query Match  
Best Local Similarity 100.0%; Score 10; DB 1; Length 448;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAG 10  
DB 159 ACCCACACAG 168

RESULT 6

ID N90574 standard; DNA; 2589 BP.  
AC N90574;  
DT 12-MAR-1990 (first entry)  
DE Caprine repeat elements GRV.1a and GRV.1b.  
KW GRV.1; Y-chromosome; ruminants; repeat element; sex determination;  
OS Capra hircus.  
PN WO8907154-A.  
PD 10-AUG-1989.  
PR 27-JAN-1989; AU00029.  
PR 29-JAN-1988; AU-006476.  
PA (ADRI-) Advanced Riverina Holdings Ltd.  
PI Reed KC; Lord EA; Matthea KI; Mann DA; Beaton S; Herr CM; Matthews ME;  
DR WPI: 89-249021/34.  
PT New nucleic acid isolates hybridising only ruminant Y chromosomal DNA -  
useful in hybridisation and polymerase chain reaction tests for sex  
PT determination.  
PS Fig. 10; 20 pp; English.  
CC The sequence is isolated from male goats and hybridises only to the DNA  
CC of cattle, sheep, goats and other ruminants which contain Y-chromosomal  
CC DNA so can be used for sex determination.  
CC Sequence 2589 BP; 754 A; 522 C; 548 G; 765 T;

Query Match  
Best Local Similarity 100.0%; Score 10; DB 1; Length 2589;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCCACACAG 10  
 DB 163 ACCCACACAG 172

## RESULT 7

ID N90766 standard; CDNA: 618 BP.

AC N90766;

DT 13-MAY-1990 (first entry)

DE Sequence of the amino acid-coding region of lymphotoxin CDNA

KW Lymphotoxin; lymphotoxin-resistant cancer cell; antitumour

FM activity.

FT Key Location/Qualifiers

FT cds 1..618

FT /\*tag= a

PD EP-314094-A.

PD 03-MAY-1989.

PD 26-OCT-1988; 117832.

PD 28-OCT-1987; JP-272034.

PD (EISA) Eisai KK.

PI Watabayashi T, Asada M, Nagasu T, Hasegawa Y, Shiketa Y,

PI Kuwada M;

PI WPI; 89-131655/18.

DR P-PSDB; P91909.

PT Recombinant lymphotoxin derivs.

PT have improved antitumour activity or exhibit a

PS cytotoxic effect on lymphotoxin-resistant cancer cells

PS Figure 1; page 38; 57pp; English.

CC A plasmid contg. CDNA coding lymphotoxin was obt'd. from a CDNA library

CC from mRNA of Daudi cells.

CC Sequence 618 BP; 110 A; 236 C; 132 G; 140 T;

Query Match 100.0%; Score 10; DB 1; Length 618;  
 Best Local Similarity 100.0%; Pred. NO. 1e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCCACACAG 10  
 DB 547 ACCCACACAG 556

## RESULT 8

ID N91654 standard; CDNA: 517 BP.

AC N91654;

DT 15-MAR-1990 (first entry)

DE CDNA sequence of tumour necrosis factor beta.

KW Tumour necrosis factor beta; CDNA.

FM Key Location/Qualifiers

FT cds 2..517

FT /\*tag= a

PD GB2217326-A.

PD 25-OCT-1989.

PD 08-APR-1988; GB-008206.

PR 08-APR-1988; 008206.

PA (BRRI-) British Bio-Technology Ltd.

PI Davies JA, Johnson ID;

PI WPI; 89-311765/43.

DR P-PSDB; P93080.

PT Synthetic gene encoding tumour necrosis factor beta - incorporates

PT restriction sites at frequent intervals to facilitate manipulation.

PS Disclosure; Fig. 1; 18pp; English.

CC The sequence is the deduced TNF beta CDNA sequence. It is prep'd.

CC synthetically.

FT Sequence 517 BP; 97 A; 199 C; 105 G; 116 T;

Query Match 100.0%; Score 10; DB 1; Length 517;  
 Best Local Similarity 100.0%; Pred. NO. 1e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCCACACAG 10  
 DB 446 ACCCACACAG 455

## RESULT 9

ID N91655 standard; DNA: 547 BP.

AC N91655;

DT 15-MAR-1990 (first entry)

DE Synthetic tumour necrosis factor beta gene.

KW Tumour necrosis factor beta.

FM Key Location/Qualifiers

FT cds 2..528

FT /\*tag= a

FT cds 535..546

FT /\*tag= b

PD GB2217326-A.

PD 25-OCT-1989.

PD 08-APR-1988; 008206.

PR 08-APR-1988; GB-008206.

PA (BRRI-) British Bio-Technology Ltd.

PI Davies JA, Johnson ID;

PI WPI; 89-311765/43.

DR P-PSDB; P93081.

PT Synthetic gene encoding tumour necrosis factor beta - incorporates

PT restriction sites at frequent intervals to facilitate manipulation.

PS Disclosure; Fig. 3; 18pp; English.

CC The synthetic TNF beta gene has the following restriction sites: HindIII,

CC BspRI, PvuII, NsiI, RsrII, KpnI, AccI, PflMI, PstI, BstXI, ScaI, BamHI

CC and EcoRI. The claimed sequence is nucleotides 14-532. The TNF beta has

CC antitumour activity. The gene can be easily modified due to the

CC restriction sites. The flanking sites simplify incorporation of the gene

CC into an expression system, eg mammalian cells, yeast and insect cell

CC hosts. The gene is produced synthetically.

CC Sequence 547 BP; 107 A; 206 C; 111 G; 123 T;

Query Match 100.0%; Score 10; DB 1; Length 547;  
 Best Local Similarity 100.0%; Pred. NO. 1e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCCACACAG 10  
 DB 461 ACCCACACAG 470

## RESULT 10

ID N91815 standard; DNA: 3086 BP.

AC N91815;

DT 19-MAR-1990 (first entry)

DE Egr-1 contg. OC3.1 clone.

KW Egr-1; DNA-binding proteins; zinc fingers;

KW early growth regulatory proteins.

OS Mus.

FM Key Location/Qualifiers

FT mat.peptide 259..1857

FT /\*tag= a

FT polyA\_signal 1865..1871

FT /\*tag= b

FT polyA\_signal 3066..3072

FT /\*tag= c

FT polyA\_signal 3053..3059

FT /\*tag= d

PD W08909777-A.

PD 19-OCT-1989.

PD 07-APR-1988; US-179587.

PR 08-APR-1988; US-249584.

PA (ARCH-) Arch Development Corp.

PI Sukhatne VP;

DR WPI; 89-324195/44.

DT	03-AUG-1990 (first entry)
DE	New synthetic human lymphotoxin (HL) gene
KW	Synthetic human lymphotoxin (HL) gene.

Job time: 18973 sec

```

FH Key      Location/Qualifiers
FT cds      CDS                      1..618
FT          /*tag= a
FT          mat_peptide      103..618
FT          /*tag= b
PN J02000447-A.
PD 05-JAN-1990.
PF 27-OCT-1987; 271513.
PR 27-OCT-1987; JP-271513, JP-287035.
PA (SANY) Sanyo KK.
PI
DR WPI: 90-047988/07.
DR P-PSDB; R05232.
PT Human lymphotoxin synthetic gene -
PS altered to leave only one restriction enzyme recognition site
PS Disclosure; Fig 1; 13pp; Japanese.
CC It is altered so that only one restriction enzyme site exists. It has
CC the same amino acid sequence but not the same base sequence as natural
CC HL and the addition of 30 nucleotides has eliminated the 6 bp palindromic
CC sequence. It makes modifications of HL easier and improved lymphotoxins
CC can be prep'd.
SQ Sequence 618 BP; 109 A; 238 C; 130 G; 141 T;

```

```

Query Match      100.0%; Score 10; DB 1; Length 618;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ACCCAGACAG 10
DB 547 ACCCAGACAG 556

```

```

RESULT 15
N81442/C
ID N81442 standard; DNA; 3344 BP.
AC N81442;
DE 17-OCT-1990 (first entry)
DE Clone encoding human protein S
KM Protein S; anti-thrombolytic agent; antibodies; ss.
OS synthetic.
FH Key      Location/Qualifiers
FT cds      247..2196
FT          /*tag= a
FT          /label=human Protein S
PN EP-255771-A.
  10-FEB-1988.
  24-JUL-1987; 306564.
  25-JUL-1986; US-890401.
PA (INTE-) Integrated Genetics.
PI Wydro R, Cohen E, Dackowski W, Stenflo J, Lundwall A, Dahlback B;
DR WPI: 88-037719/06.
DR P-PSDB; P81137.
PT Recombinant DNA encoding mature human Protein S - used therapeutically as
PT anti-thrombolytic agent and for producing antibodies for diagnostic use
PS Disclosure; P; English.
CC This clone encodes the complete mature human Protein S. To facilitate
CC expression in mammalian cells it is fused to the leader sequence of
CC bovine CDNA. See also N81441.
SQ Sequence 3344 BP; 1072 A; 575 C; 688 G; 1009 T;

```

```

Query Match      100.0%; Score 10; DB 1; Length 3344;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ACCCAGACAG 10
DB 2925 ACCCAGACAG 2916

```



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:00 ; Search time 2825.05 Seconds  
(without alignments)

6.982 Million cell updates/sec

Title: US-08-956-518a-89  
Perfect score: 10  
Sequence: 1 ACCCACACAG 10

Scoring table: IDENTITY\_NDC  
Searched: 2546578 seqs, 98626752 residues

tabase :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: em\_est20: \*  
21: em\_est21: \*  
22: em\_est22: \*  
23: em\_est23: \*  
24: em\_est24: \*  
25: em\_est25: \*  
26: em\_est26: \*  
27: em\_est27: \*  
28: em\_est28: \*  
29: em\_est29: \*  
30: em\_est30: \*  
31: em\_est31: \*  
32: em\_est32: \*  
33: em\_est33: \*  
34: em\_est34: \*  
35: em\_est35: \*  
36: em\_est36: \*  
37: em\_est37: \*  
38: em\_est38: \*  
39: em\_est39: \*  
40: em\_est40: \*  
41: em\_est41: \*  
42: em\_est42: \*  
43: em\_est43: \*  
44: em\_est44: \*  
45: em\_est45: \*  
46: em\_est46: \*  
47: em\_est47: \*  
48: em\_est48: \*  
49: em\_est49: \*  
50: em\_est50: \*  
51: em\_est51: \*  
52: em\_est52: \*  
53: em\_est53: \*

54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	10	100.0	336	20	D20150	D20150 HUMGSO1123
2	10	100.0	295	20	D22931	D22931 R1C1816A R
3	10	100.0	434	20	D24421	D24421 R1C1869A R
4	10	100.0	351	20	D24802	D24802 R1C2578A R
5	10	100.0	360	20	D36166	D36166 CELK029HAF
6	10	100.0	344	20	D39050	D39050 R1C1067A R
7	10	100.0	260	20	F01745	F01745 HSC0CB082 R
8	10	100.0	409	20	M78221	M78221 EST01814 Su
9	10	100.0	386	20	M78683	M78683 EST00831 H1
10	10	100.0	423	20	M79123	M79123 EST01271 Su
11	10	100.0	312	20	M79888	M79888 WEST00425 M
12	10	100.0	292	20	M85714	M85714 EST02233 Fe
13	10	100.0	323	20	M85791	M85791 EST02312 Fe
14	10	100.0	519	20	M88939	M88939 CEL12B2 Chr
15	10	100.0	376	20	T03144	T03144 FB25F5 Feta
16	10	100.0	475	20	T03171	T03171 FB2C12 Feta
17	10	100.0	244	20	T03346	T03346 IB1300 InfA
18	10	100.0	407	20	T03378	T03378 IB1444 InfA
19	10	100.0	451	20	T03517	T03517 IB400 InfAn
20	10	100.0	376	20	T03683	T03683 IB742 InfAn
21	10	100.0	498	20	T03711	T03711 IB781 InfAn
22	10	100.0	442	20	T03812	T03812 IB934 InfAn
23	10	100.0	440	20	T03818	T03818 IB941 InfAn
24	10	100.0	250	20	T04879	T04879 EST02767 Fe
25	10	100.0	313	20	T07793	T07793 EST05683 Fe
26	10	100.0	237	20	T08188	T08188 EST06079 In
27	10	100.0	393	20	T09172	T09172 EST07065 In
28	10	100.0	212	20	T11879	T11879 A1093R Hea
29	10	100.0	400	20	T14435	T14435 SMCST006 S
30	10	100.0	419	20	T14477	T14477 SMCST0138 S
31	10	100.0	442	20	T14830	T14830 CR8274 Lamb
32	10	100.0	454	20	T15430	T15430 IB1153 InfA
33	10	100.0	190	20	T15433	T15433 IB1181 InfA
34	10	100.0	318	20	T15490	T15490 IB1362 InfA
35	10	100.0	283	20	T15614	T15614 IB1633 InfA
36	10	100.0	271	20	T15962	T15962 IB2229 InfA
37	10	100.0	373	20	T16115	T16115 IB3508 InfA
38	10	100.0	180	20	T17312	T17312 NIB678 Norm
39	10	100.0	482	20	T23179	T23179 C2H2-604F H
40	10	100.0	319	20	T24025	T24025 seg2182 3HF
41	10	100.0	339	20	T29736	T29736 EST92350 Hu
42	10	100.0	389	20	T30661	T30661 EST20274 Hu
43	10	100.0	368	20	T31770	T31770 EST38563 Hu
44	10	100.0	318	20	T33006	T33006 EST56279 Hu
45	10	100.0	823	51	AT34960	AT34960 at05c08.x

#### ALIGNMENTS

RESULT 1  
LOCUS D20150 336 bp mRNA  
DEFINITION HUMGSO1123 Human promyelocyte Homo sapiens CDNA clone pm2161 3',  
ACCESSION D20150  
NID 9501247  
VERSION D20150.1 GI:501247

```

KEYWORDS      EST.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS        Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE          Okubo, K., Fukushima, A., Yoshii, J., Niijama, T., Kojima, Y.,
                Yoshinari, H., Arimoto, J. and Matsubara, K.
                Gene expression of human promyelocytic cell line HL60 before and
                after induction of differentiation. A new application of 3'directed
                cDNA sequencing
JOURNAL        Unpublished (1993)
COMMENT        Contact: Okubo, K., Fukushima, A., Yoshii, J., Niijama, T., Kojima, Y.,
                Yoshinari, H., Arimoto, J. and Matsubara, K.
                Institute for Molecular and Cellular Biology
                Osaka University
                3-1 Yamada-oka, Suita, Osaka 565, Japan.
FEATURES
source        1. .336
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="pm2161"
                /note="Female, adult, cell_line = HL60, cell_type =
                promyelocyte."
BASE COUNT    65 a      87 c      84 g      87 t      13 others
ORIGIN
Query Match   100.0%; Score 10; DB 20; Length 336;
Best Local Similarity 100.0%; Pred. No. 7.9e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAG 10
Db 83 ACCCACACAG 92

RESULT 2
LOCUS D22931 295 bp mRNA EST 20-JUL-1998
DEFINITION RICC1816A Rice callus Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION D22931
NID 9426856
VERSION D22931.1 GI:426856
KEYWORDS EST.
FEATURES
source        Oryza sativa.
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
                Poaceae; Oryza.
                1 (bases 1 to 295)
                Sasaki, T. and Minobe, Y.
                Rice cDNA from callus
                Unpublished (1994)
COMMENT        Contact: Takuji Sasaki
                National Institute of Agrobiological Resources
                Rice Genome Research Program
                2-1-2 Kannondai, Tsukuba
                Ibaraki,
                Japan 305
                Tel: 0298-38-7441
                Fax: 0298-38-7468
                Email: tsasaki@abrr.affrc.go.jp
                PROJECT = "RGP".
                Location/Qualifiers
                1.295
                /organism="Oryza sativa"
                /strain="cultivar Nipponbare, sub_species Japonica"
                /db_xref="taxon:4530"
                /clone="R"

FEATURES
source        1.295
                /organism="Oryza sativa"
                /strain="cultivar Nipponbare, sub_species Japonica"
                /db_xref="taxon:4530"
                /clone="R"

BASE COUNT    136 a      78 c      105 g      113 t      2 others
ORIGIN
Query Match   100.0%; Score 10; DB 20; Length 434;
Best Local Similarity 100.0%; Pred. No. 8.2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAG 10
Db 104 ACCCACACAG 95

RESULT 3
LOCUS D24421/c 434 bp mRNA EST 20-JUL-1998
DEFINITION RICR1869A Rice root Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION D24421
NID 9428273
VERSION D24421.1 GI:428273
KEYWORDS EST.
FEATURES
source        Oryza sativa.
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
                Poaceae; Oryza.
                1 (bases 1 to 434)
                Minobe, Y. and Sasaki, T.
                Rice cDNA from root
                Unpublished (1995)
COMMENT        Contact: Takuji Sasaki
                National Institute of Agrobiological Resources
                Rice Genome Research Program
                2-1-2 Kannondai, Tsukuba
                Ibaraki,
                Japan 305
                Tel: 0298-38-7441
                Fax: 0298-38-7468
                Email: tsasaki@abrr.affrc.go.jp
                PROJECT = "RGP".
                Location/Qualifiers
                1.434
                /organism="Oryza sativa"
                /strain="Nipponbare, sub_species Japonica"
                /db_xref="taxon:4530"
                /clone="R"
                /clone_lib="Rice root"
                /note="Prepared from seedling root."
BASE COUNT    50 a      89 c      112 g      37 t      7 others
ORIGIN
Query Match   100.0%; Score 10; DB 20; Length 295;
Best Local Similarity 100.0%; Pred. No. 7.8e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAG 10
Db 28 ACCCACACAG 37

RESULT 4
LOCUS D24802/c 351 bp mRNA EST 20-JUL-1998
DEFINITION RIR2578A Rice root Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION D24802
NID 9428650

```

```
/organism="Caenorhabditis elegans"  
/strain="CB1489 him-8(e1489)"
```

VERSION	F01745.1	GI:645302
NID	9645302	
COLLECTION	201703	

**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

**REFERENCE**  
 1 (bases 1 to 260)  
 Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes, M.D., Duprat, S., Houllatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Poullet, Y., Sebaste, L., Kachich, C., and Tessier, A.  
 IMAGE: molecular integration of the analysis of the human genome and its expression  
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
 95277534

**TITLE**  
 JOURNAL MEDLINE  
 95277534

**COMMENT**  
 Contact: Genethon  
 Genexpress-Genethon  
 Genethon Centre de recherche sur le Genome Humain  
 1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE  
 Tel: 33169472800  
 Fax: 33160778698  
 Email: genexpress@genethon.fr  
 Single read. removed at sequence 5' end  
 Genexpress\_library\_id: C; Genexpress\_sequence\_id: atc-0cd08  
 Seq primer: (-21)M13 universal.  
 Location/Qualifiers  
 1..260  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="17 (15)"  
 /clone="C-0cd08"  
 /clone\_lib="normalized infant brain CDNA"  
 /sex="female"  
 /tissue="total brain"  
 /dev\_stage="3 months old"  
 /note="Organ: brain; Vector: latmid BA; Site: 1: HindIII; Site: 2: NotI; sex: female; dev\_stage=3 months old; Isolate=muscular atrophy patient; tissue\_type=total brain; total mRNA was oligo-(dT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the latmid BA vector. Clone library from B.Souares, Psychiatry Dept. Columbia University, USA. Normalization\_method: Bento Soares, P.N.A.S in press"

**FEATURES**  
 Source

**BASE COUNT**  
 67 a 53 c 59 g 81 t

**ORIGIN**

Query Match  
 Best Local Similarity 100.0%; Score 10; DB 20; Length 260;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY** 1 ACCCACACAG 10  
 |||||||||  
 Db 82 ACCCACACAG 73

**RESULT** 8  
 M78221/c 409 bp mRNA EST 26-MAY-1992  
 LOCUS EST01814 Subtracted Hippocampus, Striatum (cat. #936205) Homo  
 DEFINITION sapiens cDNA clone HHCPS22 similar to Myelin basic protein, mRNA  
 sequence.  
 M78221  
 NID 9273958  
 VERSION M78221.1 GI:273958  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 Reference  
 1 (bases 1 to 409)  
 Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C.

**TITLE**  
 JOURNAL MEDLINE  
 92168112

**COMMENT**  
 Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@igf.org  
 Seq primer: M13 Forward.  
 Location/Qualifiers  
 1..409  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (1host):78662"  
 /db\_xref="GDB:D051215E"  
 /db\_xref="taxon:9606"  
 /clone="HHCPS22"  
 /clone\_lib="Subtracted Hippocampus, Striatum (cat. #936205)"  
 /note="Vector: lambdaZAP-II; The hippocampus library (#4 above) was subtracted with a fibroblast cell line cDNA library (Striatum cat. #936209; M13 lung fibroblast cell line: oligo-dT + random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb average insert size.) by the method of Sive & St. John (Nucl. Acids Res. 16:10937, 1988)."

**FEATURES**  
 Source

**BASE COUNT**  
 101 a 87 c 100 g 120 t 1 others

**ORIGIN**

Query Match  
 Best Local Similarity 100.0%; Score 10; DB 20; Length 409;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY** 1 ACCCACACAG 10  
 |||||||||  
 Db 94 ACCCACACAG 85

**RESULT** 9  
 M78683/c 386 bp mRNA EST 11-JAN-1995  
 LOCUS EST00831 Hippocampus, Striatum (cat. #936205) Homo sapiens cDNA  
 DEFINITION clone HHCPC05, mRNA sequence.  
 M78683  
 NID 9272998  
 VERSION M78683.1 GI:272998  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 Reference  
 1 (bases 1 to 386)  
 Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C.  
 Sequence identification of 2,375 human brain genes  
 Nature 355, 632-634 (1992)  
 92168112  
 2. (bases 1 to 386)  
 Durbin, A.S.  
 TITLE EST mapping data  
 JOURNAL Unpublished (1996)  
 COMMENT

Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@igf.org  
 Seq primer: M13 Reverse.

```

FEATURES
  source
    Location/Qualifiers
      1..386
        /organism="Homo sapiens"
        /db_xref="ATCC (inhost):78089"
        /db_xref="GDB:DL6S442E"
        /map="16 p13.13"
        /clone="HHC05"
        /note="Vector: Hippocampus, Stratagene (cat. #936205)"
        /note="random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb
        average insert size."
BASE COUNT      83 a      100 c      98 g      103 t      2 others
ORIGIN
Query Match      100.0%; Score 10; DB 20; Length 386;
Best Local Similarity 100.0%; Pred. No. 8e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ACCCACACAG 10
|||||
191 ACCCACACAG 182

RESULT 10
M79123      423 bp      mRNA      EST      26-MAY-1992
LOCUS      EST01271 Subtracted Hippocampus, Stratagene (cat. #936205) Homo
DEFINITION
ACCESSION   M79123
NID         927436
VERSION     M79123.1 GI:273436
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
COMMENT     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 423)
AUTHORS    Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,
            Uterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.
            Sequence identification of 2,375 human brain genes
            Nature 355, 632-634 (1992)
            92168112
TITLE      JOURNAL
JOURNAL     Nature
MEDLINE     92168112
COMMENT
  Contact: Kerlavage, AR
  Bioinformatics
  The Institute for Genomic Research
  9712 Medical Center Drive, Rockville, MD 20850 USA
  Tel: 3018699056
  Fax: 3018699423
  Email: arkerlav@tigr.org
  Seg primer: M13 Forward.
  Location/Qualifiers
    1..423
      /organism="Homo sapiens"
      /db_xref="ATCC (inhost):78667"
      /db_xref="GDB:DL6S442E"
      /db_xref="taxon:9606"
      /clone="HHC05"
      /clone.lib="Subtracted Hippocampus, Stratagene (cat.
      #936205)"
      /note="Vector: lambdaZAP-II; The hippocampus library (#4
      above) was subtracted with a fibroblast cell line cDNA
      library (Stratagene cat. #936209; W38 lung fibroblast
      cell line; oligo-dT + random primed cDNA synthesis;
      lambdaZAP-II vector, 1.0kb average insert size.) by the
      method of Sive & St. John (Nucl. Acids Res. 16:10937,
      1988)."
BASE COUNT      145 a      86 c      67 g      123 t      2 others
ORIGIN

```

```

Query Match      100.0%; Score 10; DB 20; Length 423;
Best Local Similarity 100.0%; Pred. No. 8.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ACCCACACAG 10
|||||
Db      74 ACCCACACAG 83

RESULT 11
M79888      312 bp      mRNA      EST      30-JUN-1992
LOCUS      M79888/c
DEFINITION   WEST00425 Mixed stage, Stratagene (cat. #937006) Caenorhabditis
ACCESSION   M79888
NID         9271907
VERSION     M79888.1 GI:271907
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
COMMENT     Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;
            Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
            1 (bases 1 to 312)
            McComble,W.R., Adams,M.D., Kelley,J.M., Fitzgerald,M.G.,
            Uterback,T.R., Khan,M., Dubnick,M., Kerlavage,A.R., Venter,J. and
            Fields,C.
            Caenorhabditis elegans expressed sequence tags identify gene
            families and potential disease gene homologues
            Nature Genet. 1, 124-131 (1992)
            93250983
TITLE      JOURNAL
JOURNAL     Nature
MEDLINE     93250983
COMMENT
  Contact: Kerlavage, AR
  Bioinformatics
  The Institute for Genomic Research
  9712 Medical Center Drive, Rockville, MD 20850 USA
  Tel: 3018699056
  Fax: 3018699423
  Email: arkerlav@tigr.org
  Seg primer: M13 Forward.
  Location/Qualifiers
    1..312
      /organism="Caenorhabditis elegans"
      /db_xref="taxon:6239"
      /clone="CEMSF05"
      /clone.lib="Mixed stage, Stratagene (cat. #937006)"
      /note="Vector: Uni-ZAP XR; C. elegans mixed stage cDNA
      library, Stratagene catalog #937006. The library is oligo
      dT primed and directionally cloned in the Uni-ZAP XR
      vector."
BASE COUNT      78 a      64 c      103 g      64 t      3 others
ORIGIN
Query Match      100.0%; Score 10; DB 20; Length 312;
Best Local Similarity 100.0%; Pred. No. 7.8e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ACCCACACAG 10
|||||
Db      228 ACCCACACAG 219

RESULT 12
M85714      292 bp      mRNA      EST      26-MAY-1992
LOCUS      M85714/c
DEFINITION   EST02223 Fetal brain, Stratagene (cat.#936206) Homo sapiens cDNA
ACCESSION   M85714
NID         9274361
VERSION     M85714.1 GI:274361
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens

```

REFERENCE 1 (bases 1 to 292)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 AUTHORS Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M.,  
 Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C.  
 TITLE Sequence identification of 2,375 human brain genes  
 JOURNAL Nature 355, 632-634 (1992)  
 MEDLINE 92168112  
 COMMENT

CONTACT: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org  
 Seq primer: M13 Forward.  
 Location/Qualifiers  
 1..292  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):81291"  
 /db\_xref="GDB:D0S407E"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Fetal brain, Stratagene (cat#936206)"  
 /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;  
 oligo-dT + random primed cDNA synthesis; lambdaZAP-II  
 vector, 1.0kb average insert size."  
 BASE COUNT 73 a 57 c 72 g 89 t 1 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 292;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACCCACACAG 10  
 ||||||||  
 Db 82 ACCCACACAG 73

RESULT 13  
 M85791 323 bp mRNA EST 26-MAY-1992  
 LOCUS EST02312 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA  
 DEFINITION clone HFBCN71, mRNA sequence.  
 ACCESSION M85791  
 9274440  
 M85791.1 GI:274440  
 EST.  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 323)  
 AUTHORS Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M.,  
 Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C.  
 TITLE Sequence identification of 2,375 human brain genes  
 JOURNAL Nature 355, 632-634 (1992)  
 MEDLINE 92168112  
 COMMENT

CONTACT: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org  
 Seq primer: M13 Forward.  
 Location/Qualifiers  
 1..323  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):81368"

/db\_xref="GDB:D0S482E"  
 /db\_xref="taxon:9606"  
 /clone="HFBCN71"  
 /clone\_lib="Fetal brain, Stratagene (cat#936206)"  
 /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;  
 oligo-dT + random primed cDNA synthesis; lambdaZAP-II  
 vector, 1.0kb average insert size."  
 BASE COUNT 74 a 84 c 57 g 102 t 6 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 323;  
 Best Local Similarity 100.0%; Pred. No. 7.9e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACCCACACAG 10  
 ||||||||  
 Db 25 ACCCACACAG 34

RESULT 14  
 M88939 519 bp mRNA EST 02-DEC-1992  
 LOCUS CEL12B2 Chris Martin sorted cDNA library Caenorhabditis elegans  
 DEFINITION cDNA clone cm12b2 5' similar to actin - C. elegans, mRNA sequence.  
 ACCESSION M88939  
 9275440  
 M88939.1 GI:275440  
 EST.  
 KEYWORDS Caenorhabditis elegans.  
 SOURCE Caenorhabditis elegans  
 ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;  
 Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 519)  
 AUTHORS Waterston, R., Martin, C., Craxton, M., Huynh, C., Coulson, A.,  
 Hillier, L., Durbin, R.K., Green, P., Showkeen, R., Halloran, N.,  
 Hawkins, T., Wilson, R., Berks, M., Du, Z., Thomas, K., Thierry-Mieg, J.  
 and Sulston, J.  
 TITLE A survey of expressed genes in Caenorhabditis elegans  
 JOURNAL Nature Genet. 1, 114-123 (1992)  
 MEDLINE 93250982  
 COMMENT

CONTACT: Waterston R.H. (USA) and Sulston J.E. (UK)  
 (USA) Dept. of Genetics or (UK)  
 (USA) Washington Univ. School of Medicine or (UK) MRC Laboratory of  
 Molecular Biology  
 Box 8232, 4566 Scott Ave., St. Louis, MI 63110, USA, or, Hills  
 Road, Cambridge CB2 2QH, UK  
 Tel: (USA) (314)3627072 or (UK) (0223)248011  
 Fax: (USA) (314)3624137 or (UK) (0223)402008  
 Email: rwh@genome.wustl.edu or jes@mc-lmba.cambridge.ac.uk  
 Single read.  
 Location/Qualifiers  
 1..519  
 /organism="Caenorhabditis elegans"  
 /strain="Bristol N2"  
 /db\_xref="taxon:6239"  
 /clone="cm12b2"  
 /clone\_lib="Chris Martin sorted cDNA library"  
 /lab\_host="MC1061"  
 /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;  
 oligo-dT + random primed cDNA synthesis; lambdaZAP-II  
 vector, 1.0kb average insert size."  
 BASE COUNT 121 a 162 c 106 g 128 t 2 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 519;  
 Best Local Similarity 100.0%; Pred. No. 8.3e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACCCACACAG 10  
 ||||||||  
 Db 25 ACCCACACAG 34

QY 1 ACCCAGACAG 10  
 |||||  
 Db 156 ACCCAGACAG 165

## RESULT 15

TO3144 376 bp mRNA EST 14-FEB-1997  
 LOCUS FB25F5 Fetal brain, Strata gene Homo sapiens cDNA clone FB25F5  
 DEFINITION 3' end, mRNA sequence.  
 ACCESSION TO3144  
 NID 9314385  
 VERSION TO3144.1 GI:314385  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 376)  
 Khan, A.S., Wilcox, A.S., Polymeropoulos, M.H., Hopkins, J.A.,  
 Stevens, T.J., Robinson, M., Orpana, A.K. and Sikela, J.M.

TITLE Single pass sequencing and physical and genetic mapping of human  
 CDNA

Nature Genet. 2, 180-185 (1992)

JOURNAL MEDLINE  
 COMMENT On Sep 21, 1992 this sequence version replaced gi:279283.

Contact: Sikela JM  
 Department of Pharmacology  
 University of Colorado Health Sciences Center  
 Box C236, 4200 E. 9th Ave, Denver CO 80262-0236  
 Tel: 3032708637  
 Fax: 3032707097  
 Email: mikki@italy.uchsc.edu  
 Seq primer: -21M13 Universal.  
 Location/Qualifiers

## FEATURES

## Source

1. 376  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):86789"  
 /db\_xref="GDB:D082912E"  
 /db\_xref="taxon:9606"  
 /clone="FB25F5"  
 /clone\_1b="Fetal brain, Strata gene"  
 /note="Vector: Lambda ZAP II; The FB 11 library (catalog  
 #937201, Strata gene) was constructed by directional  
 cloning and oligo d(T)-priming in lambda ZAP II phage  
 utilizing mRNA from a 17 week human fetus (total brain)."  
 101 a 90 c 73 g 109 t 3 others

-FE COUNT  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 376;  
 Best local Similarity 100.0%; Pred. No. 8e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCAGACAG 10  
 |||||  
 Db 21 ACCCAGACAG 30

Search completed: September 17, 1999, 21:28:01  
 Job time: 14265 sec





PT Human T-cell leukemia virus type-1 gag antigens -  
 PT Produced in large amounts by expression in yeast, used in  
 PT diagnostic kits or as HTLV-1 vaccine.  
 PS Disclosure: fig. 3; 15pp; English.  
 CC The gag (group specific antigen) proteins can be mass-  
 CC produced using Saccharomyces cerevisiae transformants. The  
 CC HTLV-1 gag gene or part of it can be prepared from an HTLV-1  
 CC pro virus using appropriate restriction enzymes. A large  
 CC amount of gag gene or part of it can be prepared by  
 CC amplifying in Escherichia coli a plasmid subcloned with the  
 CC HTLV-1 gag gene. DNA that encodes the gag protein is then  
 CC inserted into an expression vector. The proteins obtained  
 CC can be used in a diagnostic kit or as a HTLV-1 vaccine. See  
 CC also R03724.  
 SQ Sequence 1290 BP; 323 A; 465 C; 259 G; 243 T;

Query Match 100.0%; Score 10; DB 1; Length 1290;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CCCTATGGAG 10  
 1278 CCCTATGGAG 1269

RESULT 3  
 ID 003601 standard; DNA; 1891 BP.

AC 003601;  
 DT 02-AUG-1990 (first entry)  
 DE Gene encoding yeast aminopeptidase I.  
 KW Aminopeptidase; AP1; AP2; APX; vacuolar localisation; ds.  
 OS Saccharomyces cerevisiae.  
 FH Key Location/Qualifiers  
 FT cds 166..487  
 FT transit\_peptide 166..300  
 FT /tag= a  
 FT /label= vacuolar localisation presequence.  
 FT mat\_peptide 301..487  
 FT /tag= c  
 PN EP-359164-A.  
 PD 21-MAR-1990.  
 PE 9-SEP-1989; 116734.  
 PF 13-SEP-1988; US-243733.  
 PR 14-DEC-1988; US-284244.  
 (GEMO-) Gen Hospital Corp.  
 Smith JA, Chang Y;  
 WPI: 90-084991/12.  
 P-PSDB: R05535.

PT Substrate-specific amino:peptidase enzymes AP1, AP2 and APX -  
 PT used to remove amino-terminal amino acid residues from  
 PT peptide(s) and protein(s).  
 PS Disclosure: 24pp; English.  
 CC Aminopeptidases are useful in extending the stability, solubility and  
 CC biological activity and half-life, or in decreasing the immunogenicity.  
 CC The vacuolar location presequence encoded by the gene is useful in  
 CC directing linked polypeptides through the vacuolar membrane.  
 SQ Sequence 1891 BP; 606 A; 387 C; 388 G; 510 T;

Query Match 100.0%; Score 10; DB 1; Length 1891;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CCCTATGGAG 10  
 640 CCCTATGGAG 649

RESULT 4  
 N90123

ID N90123 standard; DNA; 5524 BP.  
 AC N90123;  
 DT 1-NOV-1989 (first entry)  
 DE DNA encoding human common acute lymphoblastic leukaemia antigen  
 KW human common acute lymphoblastic leukaemia antigen;  
 OS analgesic; inhibits leukaemia; endopeptidase.  
 FH Homo sapiens (Human)

FT cds 12  
 FT /tag= a  
 PN WO8905353-A.  
 PD 15-JUN-1989.  
 PE 01-DEC-1988; U04280.  
 PR 04-DEC-1987; US-247915.  
 PA (DANA) Dana Farber Cancer Inst.  
 PI Reinherz EL, Ship MA, Richardson NE, Ritz J, Sayre PH;  
 DR WPI: 89-192699/26.  
 P-PSDB: P90393.  
 PT DNA encoding human common acute lymphoblastic leukaemia antigen  
 PT - used for obtaining pure protein for diagnosis and  
 PT treatment of medical conditions.  
 PS Claim 4; fig 3; 73pp; English.  
 CC DNA encoding human common acute lymphoblastic leukaemia  
 CC antigen (CALA) (see corresp. P90393), and its fragments. Binds to  
 CC anti-CALA antibody, and is used for diagnosis, treatment, to  
 CC design a drug for use as a general analgesic. The DNA has functional  
 CC endopeptidase activity.  
 SQ Sequence 5524 BP; 1777 A; 1039 C; 1066 G; 1642 T;

Query Match 100.0%; Score 10; DB 1; Length 5524;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CCCTATGGAG 10  
 4546 CCCTATGGAG 4555

RESULT 5

ID N80297 standard; RNA; 4992 BP.  
 AC N80297;  
 DT 16-SEP-1990 (first entry)  
 DE Partial nucleic acid sequence of the Japanese encephalitis virus (JEV)  
 DE genome encoding C, M, E and NS proteins  
 KW Japanese encephalitis virus (JEV); vaccine; flavivirus immunity;  
 KW C protein; M protein; E protein; NS protein.  
 OS Japanese encephalitis virus.

FH Key Location/Qualifiers  
 FT mat\_peptide 1..159  
 FT /tag= a  
 FT /product= C protein  
 FT cds 160..441  
 FT /tag= b  
 FT /product= pre M protein  
 FT /tag= c  
 FT /product= M protein  
 FT /tag= d  
 FT /product= E protein  
 FT /tag= e  
 FT /product= NS1  
 FT /tag= f  
 FT /product= NS2a  
 FT /tag= g  
 FT /product= NS2b  
 FT /tag= h  
 FT mat\_peptide

```

FT      /product-NS3
PN      MO803032-A.
PD      05-MAY-1988.
PF      21-OCT-1987; 002763.
PR      27-OCT-1986; US-923907.
PA      (FOUR/) Fournier MJ.
PI      MCAda PC, Mason PM.
DR      WPI; 88-133144/19.
DR      P-PSDB; P80264.
PT      Nucleic acid of Japanese encephalitis virus -
PT      used for producing polypeptide(s) and in the diagnosis of and
PS      immunisation against the virus
PS      Claim 4: Fig 5/1-5/5: 52pp; English.
CC      Substantially purified nucleic acid comprising at least a 10 bp sequence
CC      of DNA corresponding to the nucleic acid sequence of JEV but not to the
CC      nucleic acid sequence of yellow fever virus wherein the 10 bp sequence
CC      is included within the sequence shown in n80297. Pref. the nucleic acid
CC      sequences do not correspond to those found in yellow fever virus, West
CC      Nile virus, Murray valley fever virus, dengue virus or St. Louis
CC      encephalitis virus. The nucleic acid may be obtd. from the virus and
CC      inserted into a vector. Specific probes can be derived from this
CC      nucleic acid. Also claimed are a pure polypeptide synthesized from the
CC      purified nucleic acid and antibodies produced to the polypeptide. The
CC      polypeptides and nucleic acid prods. are used in the diagnosis of and
CC      immunisation against JEV. Inoculation of animals with synthetic viral
CC      protein immunogens or with active vectors encoding such immunogens,
CC      elicits protective antiviral antibodies. The vaccination induces
CC      immunity to the disease caused by other flaviviruses.
SQ      Sequence 4992 BP; 1332 A; 1166 C; 1390 G; 1100 U; 4 Others;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 4992;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 CCTATGGAG 10
DB      1980 CCTATGGAG 1971

RESULT 6
N81434/C
ID      N81434 standard; DNA; 214 BP.
AC      N81434;
DE      11-DEC-1990 (first entry)
DE      Autonomously replicating sequence.
DE      Autonomously replicating; insulin; growth hormone; interferons; TNF;
DE      interleukins; lymphokines; enzymes; ss.
PI      Key
PI      Location/Qualifiers
FT      misc_rna
FT      80..177
FT      /tag= a
FT      /label=active fragment
FT      80..158
FT      /tag= b
FT      /label=active fragment
FT      80..105
FT      /tag= c
FT      /label=active fragment
FT      130..158
FT      /tag= d
FT      /label=active fragment
FT      150..177
FT      /tag= e
FT      /label=active fragment
FT      130..177
FT      /tag= f
FT      /label=active fragment
PN      EP-254315-A.
PD      27-JAN-1988.
PF      23-JUL-1987; 110696.
PR      24-JUL-1986; JP-174036.
PR      26-SEP-1986; JP-227455.
PA      (DAUC ) DAICHI SEIYAKU KK.

```

```

PI      Ariga H.
DR      WPI; 88-022976/04.
PT      Plasmids contg. autonomously replicating sequence DNA - useful
PT      for producing proteins
PS      Claim 9: page 10; 17pp; English.
CC      This sequence, and active fragments, is contained in an expression
CC      plasmid alongside a promoter and a gene encoding a protein of int-
CC      erest. The plasmid is propagated in host cells for the prodn. of
CC      e.g. insulin, growth hormone, TNF and lymphokines. This autonom-
CC      ously replicating sequence has an affinity for a DNA-binding prot-
CC      ein and may be recovered by binding a mammalian cell-derived DNA
CC      fragment/myc protein prod.
CC      See also N81433.
SQ      Sequence 214 BP; 37 A; 46 C; 45 G; 86 T;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 214;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 CCTATGGAG 10
DB      100 CCTATGGAG 91

RESULT 7
006691
ID      006691 standard; DNA; 2175 BP.
AC      006691;
DE      27-FEB-1991 (first entry)
DE      CDX, a MILA involved in ELAM1-mediated adhesion, from PCDM8 clone 7.2.
DE      Endothelial cell-leucocyte adhesion molecule 1; ELAM1; CDX;
DE      KW molecule involved in leucocyte adhesion; MILA; inflammation; ss.
OS      Homo sapiens.
PI      Key
PI      Location/Qualifiers
FT      cds
FT      66..1280
FT      /tag= a
FT      /product= CDX
FT      /note= "a MILA "
PN      MO9013300-A.
PD      15-NOV-1990.
PF      27-APR-1990; 002357.
PR      28-APR-1989; US-345151.
PR      01-JUN-1989; US-359516.
PR      18-DEC-1989; US-452675.
PA      (BIOG-) BIOGEN INC.
PI      Hession C, Lobb R, Goetz SE, Born I, Benjamin CD;
PI      Rose MD;
PI      WPI; 90-361248/48.
DR      P-PSDB; R08119.
PT      Endothelial cell adhesion mol. - MILAs and DNA encoding them and
PT      inhibition-detection of binding of leukocytes to endothelial
PT      cells
PS      Disclosure: Fig 9 (A-C); 136pp; English.
CC      The sequence was obtained from CDX PCDM8 clone 7.2 and from a
CC      portion of the 7.2 insert subcloned into the sequencing vector pNN11.
CC      CDX is a MILA. Involved in ELAM1-mediated adhesion and is probably
CC      the (or an) ELAM1 ligand. CDX is expressed in melanomas.
CC      See also 006686-91.
SQ      Sequence 2175 BP; 403 A; 643 C; 652 G; 477 T;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 2175;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 CCTATGGAG 10
DB      2039 CCTATGGAG 2048

RESULT 8
011140

```



Query Match 100.0%; Score 10; DB 1; Length 1591;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCCTATGAG 10  
 |||||  
 DB 29 CCCTATGAG 20

RESULT 11  
 ID 012760 standard; DNA; 3809 BP.  
 AC 012760;  
 DT 17-OCT-1991 (first entry)  
 DE P40 genomic DNA.  
 KW T cell growth factor; AIDS; cytokine; T helper cell; ss.  
 Mus musculus.  
 Key Location/Qualifiers  
 exon 644..757  
 /tag= a  
 /number= 1  
 exon 846..881  
 /tag= b  
 /number= 2  
 exon 965..997  
 /tag= c  
 /number= 3  
 exon 2184..2315  
 /tag= d  
 /number= 4  
 exon 3393..3512  
 /tag= e  
 /number= 5  
 exon 227..233  
 /tag= f  
 /label= C-rich  
 251..255  
 /tag= g  
 /label= C-rich  
 475..480  
 /tag= h  
 /note= "Interferon regulatory factor-1 binding element"  
 478..484  
 /tag= i  
 /note= "AP-1 binding site"  
 589..594  
 /tag= j  
 /tag= k  
 1361..1407  
 /tag= l  
 /label= polyd-AdC-region  
 /note= "Z DNA; enhancer activity"  
 3582..3588  
 /tag= m  
 3678..3683  
 /tag= n  
 3523..3529  
 /tag= on  
 /label= ATTATA motif  
 /note= "reduces stability of mRNA"  
 3534..3538  
 /tag= p  
 /label= ATTATA motif  
 /note= "reduces stability of mRNA"  
 3543..3547  
 /tag= q  
 /label= ATTATA motif  
 /note= "reduces stability of mRNA"  
 PN MO9110738-A.

PD 25-JUL-1991. U00145.  
 PF 08-JAN-1991; US-462158.  
 PR 08-JAN-1990; US-462158.  
 PA (RESE) RES CORP TECHN INC.  
 PI Vansnick J, Dyttenhove C, Simpson RJ;  
 DR WPI; 91-238026/32.  
 DR P-PSDB; R13218.  
 PT Nucleic acid encoding for P40 T-cell growth factor - for patients  
 PT with AIDS or compromised immune systems, also allows increased  
 PT prodn. of other cytokine(s).  
 PS Claim 1; Fig 15; 122pp; English.  
 CC The sequence was obtd. from a clone, lambda S40.1a, isolated from  
 CC a genomic library using a murine P40 cDNA clone as a probe. The  
 CC sequence, or the cDNA sequence obtd. using the sequence, can be  
 CC inserted into a vector for expression of P40 in a host organism.  
 CC The protein is useful for stimulating the proliferation of certain  
 CC subsets of T helper cells e.g. in AIDS patients or immune  
 CC compromised patients.  
 CC See also 012759 and 013660-013663.  
 SQ Sequence 3809 BP; 1013 A; 903 C; 869 G; 1024 T;

Query Match 100.0%; Score 10; DB 1; Length 3809;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCCTATGAG 10  
 |||||  
 DB 3040 CCCTATGAG 3031

RESULT 12  
 ID 014382 standard; DNA; 2175 BP.  
 AC 014382;  
 DT 13-FEB-1992 (first entry)  
 DE Clone 7.2 encoding 1,3-fucosyl transferase.  
 KW CDX; cell adhesion; ELAM1; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 66..1283  
 /tag= a  
 PN WO9116900-A.  
 PD 14-NOV-1991.  
 PF 26-OCT-1990; U06198.  
 PR 27-APR-1990; WO-002357.  
 PR 26-OCT-1990; WO-006198.  
 PA (BIOT) BIOGEN INC.  
 PI Goelz SE, Hession CA;  
 DR WPI; 91-353507/48.  
 DR P-PSDB; R14404.  
 PT DNA sequences encoding 1,3-fucosyl transferase - used to develop  
 PT anti-inflammatory therapy by inhibition of linking activity.  
 PS Claim 1; Fig 1; 38pp; English.  
 CC The clone was obtd. from CDX PCDM8 clone 7.2 and from a portion of  
 CC the 7.2 insert subcloned into the sequencing vector pNN1 to  
 CC prepare pSQ219. The sequence encodes protein 7.2, a 1,3-fucosyl  
 CC transferase a surface glycoprotein which is recognised by anti-CDX  
 CC antibodies and which binds to ELAM1. The portion of the sequence  
 CC from nucleotide 9-2162 is identical to the portion of clone 1 (Q14383)  
 CC from nucleotide 492-2645. The two inserts may represent different  
 CC transcripts from the same DNA segment. The protein may be useful in  
 CC the development of anti-inflammatory or other therapies.  
 CC See also Q14383.  
 SQ Sequence 2175 BP; 403 A; 647 C; 648 G; 477 T;

Query Match 100.0%; Score 10; DB 1; Length 2175;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCCTATGAG 10  
 |||||

Db 2039 CCCTATGAG 2048

## RESULT 13

ID Q14383 standard; DNA; 2861 BP.

AC Q14383;

DT 13-FEB-1992 (first entry)

DE Clone 1 encoding 1,3-fucosyl transferase.

KM CDX; cell adhesion; ELAM1; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT cds 74..1766

FT /tag- a

PN M09116900-A.

PD 14-NOV-1991.

PF 26-OCT-1990; WO-06198.

PR 27-APR-1990; WO-U02357.

PR 26-OCT-1990; WO-U06198.

PR (BioJ ) BIOGEN INC.

PR Goelz SE, Hession CA;

PR P-PSDB: R14405

PT DNA sequences encoding 1,3-fucosyl transferase - used to develop

PT antiinflammatory therapy by inhibition of linking activity.

PS Clain 1; Fig 2; 38pp; English.

CC The clone was obt. from CDX PCDM8 clone 1. The sequence encodes

CC protein 1, a 1,3-fucosyl transferase a surface glycoprotein which

CC is recognised by anti-CDX antibodies and which binds to ELAM1. The

CC portion of the sequence from nucleotide 492-2645 is identical to the

CC portion of clone 7.2 (Q14382) from nucleotide 9-2162. The two inserts

CC may represent different transcripts from the same DNA segment. The

CC protein may be useful in the development of anti-inflammatory or other

CC therapies.

CC See also Q14382.

SQ Sequence 2861 BP; 541 A; 840 C; 881 G; 599 T;

Query Match 100.0%; Score 10; DB 1; Length 2861;

Best Local Similarity 100.0%; Pred. No. 3.3e+02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCTATGAG 10

Db 2522 CCCTATGAG 2531

## RESULT 14

AC N50275 standard; DNA; 1530 BP.

DT 30-NOV-1991 (first entry)

DE Sequence from PATK105 contg. the p24 coding region and flanked by

DE gag gene and 5' LTR.

KM Antigen; diagnosis; assay; blood transfusion; ss.

OS Adult T-cell leukaemia virus (ATLV).

FH Key Location/Qualifiers

FT ltr 1..35

FT /tag- a

FT /tag- b

FT /product- gag

FT /tag- c

FT /product- p24

FT /tag- d

FT /product- gag

FT /tag- e

FT /tag- f

FT /tag- g

FT /tag- h

FT /tag- i

FT /tag- j

FT /tag- k

FT /tag- l

FT /tag- m

PI Taniguchi T, Yoshida M, Sugano H, Sekine S;

DR WPI: 85-106374/18.

DR P-PSDB: P50232.

PT New adult T-cell leukaemia virus antigen peptide - used in the

PT detection of T-cell leukaemia carriers in blood donors

PS Example; Table 1, Page 15-16; 37pp; English.

CC PATK105 contains a sequence extending from the 18th base pair

CC upstream from the ATG codon to the 139th base downstream from the

CC terminator codon of the ATLV gene. The example concerns the cloning

CC of the gag gene of PATK105 into pTR3 vector, which produced

CC pTAC424A which codes for a polypeptide consisting of a total of 317

CC AAs, wherein 17 AAs and 85 AAs derived from the gag gene are

CC respectively attached to the N-terminal and C-terminal of p24.

CC Escherichia coli K-12, HB101 strain which contains plasmid pTAC424A

CC has been deposited with the Fermentation Research Institute as

CC Escherichia coli ETAC424A, FERM BP-341.

SQ Sequence 1350 BP; 386 A; 546 C; 306 G; 292 T;

Query Match 100.0%; Score 10; DB 1; Length 1350;

Best Local Similarity 100.0%; Pred. No. 3.2e+02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCTATGAG 10

Db 1359 CCCTATGAG 1350

## RESULT 15

ID N40080 standard; cDNA; 9047 BP.

AC N40080;

DT 04-FEB-1992 (first entry)

DE Sequence complementary to the genome of adult T-cell leukaemia virus

DE (ATLV) ATK-1.

KM Diagnosis; therapy; prevention; leukaemia; antigen; ss.

OS Adult T-cell leukaemia virus.

FH Key Location/Qualifiers

FT 5'utr 1..755

FT /tag- a

FT /tag- b

FT /product- gag

FT /tag- c

FT /product- pol

FT /tag- d

FT /product- env

FT /tag- e

FT /product- px-I

FT /tag- f

FT /product- px-II

FT /tag- g

FT /product- px-III

FT /tag- h

FT /product- px-IV

FT /tag- i

FT /product- px-V

FT /tag- j

FT /product- px-VI

FT /tag- k

FT /product- px-VII

FT /tag- l

FT /product- px-VIII

FT /tag- m

FT /product- px-IX

FT /tag- n

FT /product- px-X

FT /tag- o

FT /product- px-XI

FT /tag- p

FT /product- px-XII

PI Taniguchi T, Yoshida M, Sugano H, Sekine S;

DR WPI: 85-106374/18.

DR P-PSDB: P50232.

PT New adult T-cell leukaemia virus antigen peptide - used in the

PT detection of T-cell leukaemia carriers in blood donors

PS Example; Table 1, Page 15-16; 37pp; English.

CC PATK105 contains a sequence extending from the 18th base pair

CC upstream from the ATG codon to the 139th base downstream from the

CC terminator codon of the ATLV gene. The example concerns the cloning

CC of the gag gene of PATK105 into pTR3 vector, which produced

CC pTAC424A which codes for a polypeptide consisting of a total of 317

CC AAs, wherein 17 AAs and 85 AAs derived from the gag gene are

CC respectively attached to the N-terminal and C-terminal of p24.

CC Escherichia coli K-12, HB101 strain which contains plasmid pTAC424A

CC has been deposited with the Fermentation Research Institute as

CC Escherichia coli ETAC424A, FERM BP-341.

SQ Sequence 1350 BP; 386 A; 546 C; 306 G; 292 T;

Query Match 100.0%; Score 10; DB 1; Length 1350;

Best Local Similarity 100.0%; Pred. No. 3.2e+02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCTATGAG 10

Db 1359 CCCTATGAG 1350

## RESULT 15

ID N40080 standard; cDNA; 9047 BP.

AC N40080;

DT 04-FEB-1992 (first entry)

DE Sequence complementary to the genome of adult T-cell leukaemia virus

DE (ATLV) ATK-1.

KM Diagnosis; therapy; prevention; leukaemia; antigen; ss.

OS Adult T-cell leukaemia virus.

FH Key Location/Qualifiers

FT 5'utr 1..755

FT /tag- a

FT /tag- b

FT /product- gag

FT /tag- c

FT /product- pol

FT /tag- d

FT /product- env

FT /tag- e

FT /product- px-I

FT /tag- f

FT /product- px-II

FT /tag- g

FT /product- px-III

FT /tag- h

FT /product- px-IV

FT /tag- i

FT /product- px-V

FT /tag- j

FT /product- px-VI

FT /tag- k

FT /product- px-VII

FT /tag- l

FT /product- px-VIII

FT /tag- m

FT /product- px-IX

FT /tag- n

FT /product- px-X

FT /tag- o

FT /product- px-XI

FT /tag- p

FT /product- px-XII

PI Taniguchi T, Yoshida M, Sugano H, Sekine S;

DR WPI: 85-106374/18.

DR P-PSDB: P50232.

PT New adult T-cell leukaemia virus antigen peptide - used in the

PT detection of T-cell leukaemia carriers in blood donors

PS Example; Table 1, Page 15-16; 37pp; English.

CC PATK105 contains a sequence extending from the 18th base pair

CC upstream from the ATG codon to the 139th base downstream from the

CC terminator codon of the ATLV gene. The example concerns the cloning

CC of the gag gene of PATK105 into pTR3 vector, which produced

CC pTAC424A which codes for a polypeptide consisting of a total of 317

CC AAs, wherein 17 AAs and 85 AAs derived from the gag gene are

CC respectively attached to the N-terminal and C-terminal of p24.

CC Escherichia coli K-12, HB101 strain which contains plasmid pTAC424A

CC has been deposited with the Fermentation Research Institute as

CC Escherichia coli ETAC424A, FERM BP-341.

SQ Sequence 1350 BP; 386 A; 546 C; 306 G; 292 T;

Query Match 100.0%; Score 10; DB 1; Length 1350;

Best Local Similarity 100.0%; Pred. No. 3.2e+02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCTATGAG 10

Db 1359 CCCTATGAG 1350

## RESULT 15

ID N40080 standard; cDNA; 9047 BP.

AC N40080;

DT 04-FEB-1992 (first entry)

DE Sequence complementary to the genome of adult T-cell leukaemia virus

DE (ATLV) ATK-1.

KM Diagnosis; therapy; prevention; leukaemia; antigen; ss.

OS Adult T-cell leukaemia virus.

FH Key Location/Qualifiers

FT 5'utr 1..755

FT /tag- a

FT /tag- b

FT /product- gag

FT /tag- c

FT /product- pol

FT /tag- d

FT /product- env

FT /tag- e

FT /product- px-I

FT /tag- f

FT /product- px-II

FT /tag- g

FT /product- px-III

FT /tag- h

FT /product- px-IV

FT /tag- i

FT /product- px-V

FT /tag- j

FT /product- px-VI

FT /tag- k

FT /product- px-VII

FT /tag- l

FT /product- px-VIII

FT /tag- m

FT /product- px-IX

FT /tag- n

FT /product- px-X

FT /tag- o

FT /product- px-XI

FT /tag- p

FT /product- px-XII

PI Taniguchi T, Yoshida M, Sugano H, Sekine S;

DR WPI: 85-106374/18.

DR P-PSDB: P50232.

PT New adult T-cell leukaemia virus antigen peptide - used in the

PT detection of T-cell leukaemia carriers in blood donors

PS Example; Table 1, Page 15-16; 37pp; English.

CC PATK105 contains a sequence extending from the 18th base pair

CC upstream from the ATG codon to the 139th base downstream from the

CC terminator codon of the ATLV gene. The example concerns the cloning

CC of the gag gene of PATK105 into pTR3 vector, which produced

CC pTAC424A which codes for a polypeptide consisting of a total of 317

CC AAs, wherein 17 AAs and 85 AAs derived from the gag gene are

CC respectively attached to the N-terminal and C-terminal of p24.

CC Escherichia coli K-12, HB101 strain which contains plasmid pTAC424A

CC has been deposited with the Fermentation Research Institute as

CC Escherichia coli ETAC424A, FERM BP-341.

SQ Sequence 1350 BP; 386 A; 546 C; 306 G; 292 T;

Query Match 100.0%; Score 10; DB 1; Length 1350;

Best Local Similarity 100.0%; Pred. No. 3.2e+02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCTATGAG 10

Db 1359 CCCTATGAG 1350

## RESULT 15

ID N40080 standard; cDNA; 9047 BP.

AC N40080;

DT 04-FEB-1992 (first entry)

DE Sequence complementary to the genome of adult T-cell leukaemia virus

DE (ATLV) ATK-1.

KM Diagnosis; therapy; prevention; leukaemia; antigen; ss.

OS Adult T-cell leukaemia virus.

FH Key Location/Qualifiers

FT 5'utr 1..755

FT /tag- a

FT /tag- b

FT /product- gag

FT /tag- c

FT /product- pol

FT /tag- d

FT /product- env

FT /tag- e

FT /product- px-I

FT /tag- f

FT /product- px-II

FT /tag- g

FT /product- px-III

FT /tag- h

FT /product- px-IV

PI Yoshida M, Sugano H;  
DR WPI; 84-172336/28.  
PT Viral genomic DNA complementary to RNA of human leukaemia virus -  
PT useful in recombinant DNA producing therapeutic and diagnostic  
PT proteins  
PS Claim 4; Table 1, Page 11-15; 23pp; English.  
CC Diagnosis of human leukaemia and/or lymphoma and virus infection is  
CC made with part or all of the recombinant DNA. Virus antigenic  
CC proteins can be produced. These peptides and proteins, and  
CC antibodies against them, are useful for the diagnosis, therapy and  
CC prevention of human leukaemia.  
SQ Sequence 9047 BP; 2087 A; 3164 C; 1713 G; 2083 T;

Query Match 100.0%; Score 10; DB 1; Length 9047;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CCCTATGAG 10  
|||||||  
2079 CCCTATGAG 2070

Search completed: September 18, 1999, 00:33:35  
Job time: 18975 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:01 : Search time 2825.05 Seconds

(without alignments)  
6.962 Million cell updates/sec

Title: US-08-956-518a-90  
Perfect score: 10  
Sequence: 1 CCCTATGAG 10

Scoring table: IDENTITY\_NUC

arched: 2546578 seqs, 98626752 residues

-cbase :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	10	100.0	257	20	D19706	D19706 HUMG00662
2	10	100.0	360	20	D35906	D35906 CEIK02685F
3	10	100.0	360	20	D35977	D35977 CEIK02777F
4	10	100.0	298	20	D39999	D39999 RICS1711A R
5	10	100.0	188	20	F01533	F01533 HSC05H042 n
6	10	100.0	312	20	F03038	F03038 HSC1KD092 n
7	10	100.0	349	20	M85318	M85318 EST01434 Fe
8	10	100.0	335	20	T01239	T01239 WST01960 E
9	10	100.0	667	20	T03839	T03839 GT0024 Gene
10	10	100.0	422	20	T03843	T03843 GT0131 Gene
11	10	100.0	236	20	T03944	T03944 GT1159 Gene
12	10	100.0	400	20	T04417	T04417 464 Lambda-
13	10	100.0	413	20	T04785	T04785 833 Lambda-
14	10	100.0	357	20	T06534	T06534 EST04423 Fe
15	10	100.0	453	20	T13875	T13875 2040 Lambda
16	10	100.0	389	20	T18860	T18860 B090205 Tes
17	10	100.0	149	20	T25900	T25900 ATH014 HPCD
18	10	100.0	339	20	T26587	T26587 AB327H2F In
19	10	100.0	264	20	T27698	T27698 EST12669 Hu
20	10	100.0	223	20	T29507	T29507 EST82429 Hu
21	10	100.0	210	20	T31409	T31409 EST31977 Hu
22	10	100.0	306	20	T31520	T31520 EST34096 Hu
23	10	100.0	329	20	T31768	T31768 EST38526 Hu
24	10	100.0	348	20	T32482	T32482 EST49297 Hu
25	10	100.0	297	20	T32537	T32537 EST50461 Hu
26	10	100.0	307	20	T33954	T33954 EST60289 Hu
27	10	100.0	381	20	T34070	T34070 EST62150 Hu
28	10	100.0	319	20	T35772	T35772 EST91174 Hu
29	10	100.0	482	20	T50468	T50468 YB30G01.s1
30	10	100.0	372	20	T52181	T52181 YB2811.s1
31	10	100.0	430	20	T54450	T54450 YB06H04.r2
32	10	100.0	237	20	Z22575	Z22575 M142/1AB In
33	10	100.0	320	20	Z25513	Z25513 ATTS1026 Gr
34	10	100.0	168	20	Z31029	Z31029 NMTEST128 M
35	10	100.0	412	20	Z37642	Z37642 ATTS4107 Ve
36	10	100.0	317	20	Z40389	Z40389 HSC21B082 n
37	10	100.0	337	20	Z42738	Z42738 HSC0SE101 n
38	10	100.0	330	20	Z42925	Z42925 HSC0FA121 n
39	10	100.0	282	20	Z44843	Z44843 HSC2AG121 n
40	10	100.0	206	20	Z45940	Z45940 HSC2XG121 n
41	10	100.0	308	20	Z46007	Z46007 HSC0BP021 n
42	10	100.0	258	21	F05474	F05474 HSC0CC011 n
43	10	100.0	313	21	F05489	F05489 HSC0CF091 n
44	10	100.0	359	21	F06540	F06540 HSC1BE071 n
45	10	100.0	590	51	A1734774	A1734774 606036C04

#### ALIGNMENTS

RESULT 1  
LOCUS D19706/c  
DEFINITION HUMG00662 Human promyelocyte Homo sapiens  
ACCESSION D19706  
NID 9500022  
VERSION D19706.1 GI:500022

EST 30-JUL-1996  
clone mm08f12 3',

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 257)  
AUTHORS Okubo, K., Fukushima, A., Yoshii, J., Niijima, T., Kojima, Y.,  
Yoshinari, H., Arimoto, J. and Matsubara, K.  
TITLE Gene expression of human promyelocytic cell line HL60 before and  
after induction of differentiation. A new application of 3'directed  
cDNA sequencing  
JOURNAL Unpublished (1993)  
COMMENT Contact: Okubo, K., Fukushima, A., Yoshii, J., Niijima, T., Kojima, Y.,  
Yoshinari, H., Arimoto, J. and Matsubara, K.  
Institute for Molecular and Cellular Biology  
Osaka University  
3-1 Yamada-oka, Suita, Osaka 565, Japan.  
Location/Qualifiers  
FEATURES  
source 1..257  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Human promyelocyte"  
/note="Female, adult, cell\_line = HL60, cell\_type =  
promyelocyte." 54 g 79 t 5 others  
BASE COUNT 84 a 35 c  
ORIGIN  
Query Match 100.0%; Score 10; DB 20; Length 257;  
Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCTATGAG 10  
|||||  
Db 174 CCTATGAG 165  
RESULT 2  
LOCUS D35906 360 bp mRNA EST 08-AUG-1994  
DEFINITION CELK0265F Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA  
clone YK26h5 5', mRNA sequence.  
ACCESSION D35906  
NID 9527845  
VERSION D35906.1 GI:527845  
KEYWORDS EST.  
RCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;  
Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
REFERENCE 1 (bases 1 to 360)  
AUTHORS Kohara, Y., Mitsuiki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and  
Tabara, H.  
TITLE Toward an expression map of the C.elegans genome  
JOURNAL Unpublished (1994)  
COMMENT Contact: Yuji Kohara  
Gene Library Lab  
National Institute of Genetics  
Yata 1111, Mishima, Shizuoka 411, Japan  
Tel: 0559-75-0771  
Fax: 0559-75-6240  
Email: ykohara@dbj.nig.ac.jp.  
Location/Qualifiers  
FEATURES  
source 1..360  
/organism="Caenorhabditis elegans"  
/strain="CB1489 him-8(e1489)"  
/db\_xref="taxon:6239"  
/clone\_lib="Yuji Kohara unpublished cDNA"  
/note="dev\_stage=varied, sex=Hermaphrodite male,  
tissue\_type=whole animal." 81 g 102 t 7 others  
BASE COUNT 92 a 78 c  
ORIGIN  
Query Match 100.0%; Score 10; DB 20; Length 360;  
Best Local Similarity 100.0%; Pred. No. 2e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCTATGAG 10  
|||||  
Db 21 CCTATGAG 30  
RESULT 4  
LOCUS D39999 298 bp mRNA EST 11-NOV-1994  
DEFINITION RICS1711A Rice shoot Oryza sativa cDNA, mRNA sequence.  
ACCESSION D39999  
NID 9569150  
VERSION D39999.1 GI:569150  
KEYWORDS EST.  
SOURCE Oryza sativa.  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

BASE COUNT 104 a 66 c 83 g 103 t 4 others  
ORIGIN  
Query Match 100.0%; Score 10; DB 20; Length 360;  
Best Local Similarity 100.0%; Pred. No. 2e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCTATGAG 10  
|||||  
Db 154 CCTATGAG 163  
RESULT 3  
LOCUS D35977 360 bp mRNA EST 08-AUG-1994  
DEFINITION CELK027F7F Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA  
clone YK27f7 5', mRNA sequence.  
ACCESSION D35977  
NID 9527915  
VERSION D35977.1 GI:527915  
KEYWORDS EST.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;  
Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
REFERENCE 1 (bases 1 to 360)  
AUTHORS Kohara, Y., Mitsuiki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and  
Tabara, H.  
TITLE Toward an expression map of the C.elegans genome  
JOURNAL Unpublished (1994)  
COMMENT Contact: Yuji Kohara  
Gene Library Lab  
National Institute of Genetics  
Yata 1111, Mishima, Shizuoka 411, Japan  
Tel: 0559-75-0771  
Fax: 0559-75-6240  
Email: ykohara@dbj.nig.ac.jp.  
Location/Qualifiers  
FEATURES  
source 1..360  
/organism="Caenorhabditis elegans"  
/strain="CB1489 him-8(e1489)"  
/db\_xref="taxon:6239"  
/clone\_lib="Yuji Kohara unpublished cDNA"  
/note="dev\_stage=varied, sex=Hermaphrodite male,  
tissue\_type=whole animal." 81 g 102 t 7 others  
BASE COUNT 92 a 78 c 83 g 103 t 4 others  
ORIGIN  
Query Match 100.0%; Score 10; DB 20; Length 360;  
Best Local Similarity 100.0%; Pred. No. 2e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCTATGAG 10  
|||||  
Db 21 CCTATGAG 30  
RESULT 4  
LOCUS D39999 298 bp mRNA EST 11-NOV-1994  
DEFINITION RICS1711A Rice shoot Oryza sativa cDNA, mRNA sequence.  
ACCESSION D39999  
NID 9569150  
VERSION D39999.1 GI:569150  
KEYWORDS EST.  
SOURCE Oryza sativa.  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;



euphylliphytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
Poaceae; Oryza.  
1 (bases 1 to 298)  
Sasaki,T., Miyao,A. and Yamamoto,K.  
Rice cDNA from callus 1995  
Unpublished (1995)

## COMMENT

Contact: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program  
2-1-2 Kannondai, Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441  
Fax: 0298-38-7468  
Email: tsasaki@abr.affrc.go.jp.

## FEATURES

Location/Qualifiers  
1..298  
/organism="Oryza sativa"  
/strain="Nipponbare, sub\_species Japonica"  
/db\_xref="taxon:4530"  
/map="8"  
/clone\_lib="Rice shoot"  
/note="Etolated shoot (8 days old)"  
BASE COUNT 91 a 61 c 62 g 79 t 5 others  
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 298;  
Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CCTATGGAG 10  
|||||  
Db 107 CCTATGGAG 116

## RESULT 5

F01533/c 188 bp mRNA EST 01-FEB-1995  
LOCUS HSC05H042 normalized Infant brain cDNA Homo sapiens cDNA clone  
DEFINITION c-05H04 3', mRNA sequence.

## ACCESSION

F01533  
9645090  
F01533.1 GI:645090

## VERSION

EST.

## WORDS

human.

## RCE

human.

## ORGANISM

Homo sapiens

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 188)

## AUTHORS

Aufiray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,  
Devignes,M.D., Duprat,S., Houligatte,R., Juneau,M.N., Lamy,B.,  
Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pletu,G., Pouillot,Y.,  
Sebastiani-Kabaktchis,C. and Tessier,A.  
IMAGE: molecular integration of the analysis of the human genome  
and its expression

## TITLE

C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)

## JOURNAL

95277534

## MEDLINE

On Oct 24, 1994 this sequence version replaced.

## COMMENT

Contact: Genethon  
Genexpress-Genethon  
Genethon Centre de recherche sur le Genome Humain  
1,rue de l'Internationale, Bp60 91002 EVRY Cedex, FRANCE  
Tel: 33169472800  
Fax: 33160778698  
Email: genexpress@genethon.fr  
Single read. removed at sequence 5'end  
Genexpress\_library\_idt: C; Genexpress\_sequence\_idt: alc-05h04  
Seq primer: (-21)M13-universal.

## FEATURES

Location/Qualifiers  
1..188

Source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="c-05H04"  
/clone\_lib="normalized Infant brain cDNA"  
/sex="Female"  
/tissue\_type="total brain"  
/dev\_stage="3 months old"  
/note="Organ: brain; Vector: lafmd BA; Site\_1: HindIII;  
Site\_2: NotI; sex:female; dev\_stage=3 months old;  
Isolate=muscular atrophy patient; tissue\_type=total  
brain; total mRNA was oligo-(dT) primed and directionally  
cloned 5' -> 3' into the HindIII -> NotI sites of the  
lafmd BA vector. Clone library from B Soares, Psychiatry  
Dept. Columbia University, USA. Normalization method:  
Bento Soares, P.N.A.S in press"

BASE COUNT 59 a 48 c 30 g 47 t 4 others  
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 188;  
Best Local Similarity 100.0%; Pred. No. 1.8e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CCTATGGAG 10  
|||||  
Db 156 CCTATGGAG 147

## RESULT 6

F03038 312 bp mRNA EST 02-FEB-1995  
LOCUS HSC1KD092 normalized Infant brain cDNA Homo sapiens cDNA clone  
DEFINITION c-1kd09 3', mRNA sequence.

## ACCESSION

F03038  
9646595  
F03038.1 GI:646595

## VERSION

EST.

## KEYWORDS

human.

## SOURCE

Homo sapiens

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 312)

## AUTHORS

Aufiray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,  
Devignes,M.D., Duprat,S., Houligatte,R., Juneau,M.N., Lamy,B.,  
Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pletu,G., Pouillot,Y.,  
Sebastiani-Kabaktchis,C. and Tessier,A.  
IMAGE: molecular integration of the analysis of the human genome  
and its expression

## TITLE

C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)

## JOURNAL

95277534

## MEDLINE

Contact: Genethon  
Genexpress-Genethon  
Genethon Centre de recherche sur le Genome Humain  
1,rue de l'Internationale, Bp60 91002 EVRY Cedex, FRANCE  
Tel: 33169472800  
Fax: 33160778698  
Email: genexpress@genethon.fr  
Single read. removed at sequence 5'end  
Genexpress\_library\_idt: C; Genexpress\_sequence\_idt: alc-1kd09  
Seq primer: (-21)M13-universal.

## COMMENT

Location/Qualifiers  
1..312

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="1"  
/clone="c-1kd09"  
/clone\_lib="normalized Infant brain cDNA"  
/sex="Female"  
/tissue\_type="total brain"  
/dev\_stage="3 months old"  
/note="Organ: brain; Vector: lafmd BA; Site\_1: HindIII;  
Site\_2: NotI; sex:female; dev\_stage=3 months old"

## FEATURES

Source

isolate-muscular atrophy patient; tissue-type-total  
 brain; total mRNA was oligo-(GT) primed and directionally  
 cloned 5' -> 3' into the HindIII -> NotI sites of the  
 lambdaB vector. Clone library from B. Soares, Psychiatry  
 Dept. Columbia University, USA. Normalization method:  
 Bento Soares, P.N.A.S. in press"

BASE COUNT 91 a 72 c 55 g 90 t 4 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 312;  
 Best Local Similarity 100.0%; Pred. No. 2e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCCATGAG 10  
 |||||

DB 155 CCCATGAG 146

RESULT 7  
 118  
 US M85318 349 bp mRNA EST 26-MAY-1992  
 DEFINITION EST01434 Fetal brain, Strata gene (cat#936206) Homo sapiens CDNA  
 clone HFBBA18 similar to Alu repetitive element, mRNA sequence.  
 ACCESSION M85318  
 NID 9273588  
 VERSION M85318.1 GI:273588  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 349)  
 Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,  
 Uterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.  
 Sequence identification of 2,375 human brain genes  
 JOURNAL Nature 355, 632-634 (1992)  
 MEDLINE 92168112  
 COMMENT

Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlavage@tigr.org  
 Seq primer: M13 Forward.

FEATURES  
 source  
 Location/Qualifiers  
 1..349  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):77543"  
 /db\_xref="GDB:DO511E"  
 /db\_xref="taxon:9606"  
 /clone="HFBBA18"  
 /clone\_idb="Fetal brain, Strata gene (cat#936206)"  
 /note="vector: LambdaZAP-II; 17-18 wk gestation, female;  
 oligo-dT + random primed cDNA synthesis; lambdaZAP-II  
 vector, 1.0kb average insert size."  
 72 a 96 c 73 g 101 t 7 others

BASE COUNT

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 349;  
 Best Local Similarity 100.0%; Pred. No. 2e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCCATGAG 10  
 |||||

DB 239 CCCATGAG 248

RESULT 8  
 T01239/c

LOCUS T01239 335 bp mRNA EST 10-NOV-1992  
 DEFINITION WEST01960 Early embryo, Strata gene (cat. #937007) Caenorhabditis  
 elegans CDNA clone CEES044, mRNA sequence.  
 ACCESSION T01239  
 NID 9277720  
 VERSION T01239.1 GI:277720  
 KEYWORDS EST.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderae; Caenorhabditis.  
 REFERENCE 1 (bases 1 to 335)  
 McCombie,W.R., Kelley,J.M., Aubin,L., Goscoechea,M.,  
 FitzGerald,M.G., Wu,A., Adams,M.D., Dubnick,M., Kerlavage,A.R.,  
 Venter,J.C. and Fields,C.A.  
 Caenorhabditis elegans cDNAs  
 JOURNAL Unpublished (1993)  
 MEDLINE

COMMENT

Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlavage@tigr.org  
 Seq primer: M13 Forward.

FEATURES  
 source  
 Location/Qualifiers  
 1..335  
 /organism="Caenorhabditis elegans"  
 /db\_xref="taxon:6239"  
 /clone="CEES044"  
 /clone\_idb="Early embryo, Strata gene (cat. #937007)"

BASE COUNT 116 a 67 c 64 g 85 t 3 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 335;  
 Best Local Similarity 100.0%; Pred. No. 2e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCCATGAG 10  
 |||||

DB 309 CCCATGAG 300

RESULT 9  
 T03839/c

LOCUS T03839 667 bp mRNA EST 14-JUL-1993  
 DEFINITION GT0024 Genetrack, 4p16.3 JM Rommens Homo sapiens CDNA clone gt24  
 similar to some l1 sequence similarity, mRNA sequence.  
 ACCESSION T03839  
 NID 9315075  
 VERSION T03839.1 GI:315075  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 667)  
 Rommens,J.M., Lin,B., Hutchinson,G.B., Andrew,S.E., Goldberg,Y.P.,  
 Glaves,M.L., Graham,R., Lai,V., McArthur,J., Nasir,J. et al.  
 A transcription map of the region containing the Huntington disease  
 gene  
 JOURNAL Hum. Mol. Genet. 2, 901-907 (1993)  
 MEDLINE 93372851

COMMENT

Contact: Hutchinson GB  
 Dept. of Medical Genetics, University of British Columbia  
 Canadian Genetic Diseases Network  
 Rm. 416, 2125 East Mall, Vancouver, B.C. V6T 1Z4  
 Tel: (604)822-9240  
 Fax: (604)822-9238  
 Email: hutch@uam.genetics.ca.

FEATURES  
SOURCE

Location/Qualifiers  
1..667  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="6 p23-p22"  
/clone="gt24"

BASE COUNT  
ORIGIN

/note="CDNA from frontal cortex RNA and from a pool prepared from fetal brain, frontal cortex, bone marrow and liver RNA. The cDNAs were hybridized to immobilized isolated yeast artificial chromosomes from the 4p16.3 region. Following washing at high stringency, cDNA segments that had specifically hybridized were eluted, amplified by PCR and cloned. Clones were mapped back to the YACs. Inserts were sequenced using both standard (T3 and T7) and custom primers."

187 a 120 c 98 g 262 t

Query Match  
Best Local Similarity 100.0%; Score 10; DB 20; Length 667;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CCCTATGGAG 10  
|||||  
Db 203 CCCTATGGAG 194

## RESULT 10

LOCUS T03843 422 bp mRNA EST 14-JUL-1993  
DEFINITION GT0131 Genetrack, 4p16.3 JM Rommens Homo sapiens CDNA clone gt131,  
mRNA sequence.

## ACCESSION

T03843

## NID

9315082

## VERSION

T03843.1 GI:315082

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

## REFERENCE

1. (bases 1 to 422)  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

## AUTHORS

Rommens,J.M., Lin,B., Hutchinson,G.B., Andrew,S.E., Goldberg,Y.P.,  
Glaves,M.L., Graham,R., Lai,V., McArthur,J., Nasir,J. et al.

## TITLE

A transcription map of the region containing the Huntington disease  
gene

## JOURNAL

Hum. Mol. Genet. 2, 901-907 (1993)

## MEDLINE

93372851

## COMMENT

On Apr 14, 1993 this sequence version replaced.

## FEATURES

## SOURCE

Contact: Hutchinson GB  
Dept. of Medical Genetics, University of British Columbia  
Canadian Genetic Diseases Network  
Rm. 416, 2125 East Mall, Vancouver, B.C. V6T 1Z4  
Tel: (604)822-9240  
Fax: (604)822-9238  
Email: hutch@hug.med.ubc.ca.  
Location/Qualifiers  
1..422  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="gt131"  
/note="CDNA from frontal cortex RNA and from a pool prepared from fetal brain, frontal cortex, bone marrow and liver RNA. The cDNAs were hybridized to immobilized isolated yeast artificial chromosomes from the 4p16.3 region. Following washing at high stringency, cDNA segments that had specifically hybridized were eluted, amplified by PCR and cloned. Clones were mapped back to the YACs. Inserts were sequenced using both standard (T3 and T7) and custom primers."

BASE COUNT 157 a 108 c 107 g 50 t  
ORIGIN

Query Match  
Best Local Similarity 100.0%; Score 10; DB 20; Length 422;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CCCTATGGAG 10  
|||||  
Db 259 CCCTATGGAG 268

## RESULT 11

LOCUS T03944 236 bp mRNA EST 14-JUL-1993  
DEFINITION GT1159 Genetrack, 4p16.3 JM Rommens Homo sapiens CDNA clone gt159y,  
mRNA sequence.

## ACCESSION

T03944

## NID

9315102

## VERSION

T03944.1 GI:315102

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

## REFERENCE

1. (bases 1 to 236)  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

## AUTHORS

Rommens,J.M., Lin,B., Hutchinson,G.B., Andrew,S.E., Goldberg,Y.P.,  
Glaves,M.L., Graham,R., Lai,V., McArthur,J., Nasir,J. et al.

## TITLE

A transcription map of the region containing the Huntington disease  
gene

## JOURNAL

Hum. Mol. Genet. 2, 901-907 (1993)

## MEDLINE

93372851

## COMMENT

On Apr 14, 1993 this sequence version replaced.

## FEATURES

## SOURCE

Contact: Hutchinson GB  
Dept. of Medical Genetics, University of British Columbia  
Canadian Genetic Diseases Network  
Rm. 416, 2125 East Mall, Vancouver, B.C. V6T 1Z4  
Tel: (604)822-9240  
Fax: (604)822-9238  
Email: hutch@hug.med.ubc.ca.  
Location/Qualifiers  
1..236  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="gt159y"  
/note="CDNA from frontal cortex RNA and from a pool prepared from fetal brain, frontal cortex, bone marrow and liver RNA. The cDNAs were hybridized to immobilized isolated yeast artificial chromosomes from the 4p16.3 region. Following washing at high stringency, cDNA segments that had specifically hybridized were eluted, amplified by PCR and cloned. Clones were mapped back to the YACs. Inserts were sequenced using both standard (T3 and T7) and custom primers."

BASE COUNT 39 a 48 c 54 g 88 t 7 others  
ORIGIN

Query Match  
Best Local Similarity 100.0%; Score 10; DB 20; Length 236;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CCCTATGGAG 10  
|||||  
Db 162 CCCTATGGAG 153

## RESULT 12

LOCUS T04417 400 bp mRNA EST 06-NOV-1997  
DEFINITION 464 Lambda-PRL2 Arabidopsis thaliana CDNA clone 40E11T7, mRNA

sequence.  
 accession 104417  
 NID 9315577  
 VERSION 104417.1 GI:315577  
 KEYWORDS EST.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 400)  
 REFERENCE Newman,T., deBrujn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M., Retzel,E. and Somerville,C.  
 TITLE Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones Plant Physiol. 106, 1241-1255 (1994)  
 JOURNAL MEDLINE 95148729  
 COMMENT  
 CONTACT: Thomas Newman  
 MSU-DOE Plant Research Laboratory  
 Michigan State University  
 MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E. Lansing, MI  
 Tel: 517-353-0854  
 Fax: 517-353-9168  
 Email: 22313cne@dm.cl.msu.edu  
 Seq primer: T7 dye primer.  
 Location/Qualifiers  
 1..400  
 /organism="Arabidopsis thaliana"  
 /strain="var columbia"  
 /db\_xref="taxon:3702"  
 /clone="40E117"  
 /clone\_1db="lambda-PRL2"  
 /note="Vector: lambda Zip-lox; Site\_1: Sal; Site\_2: Not; Lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques). The vector is BRV's lambda Zip-lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dt primed cDNA."  
 BASE COUNT 92 a 131 c 62 g 113 t 2 others  
 ORIGIN  
 Query Match 100.0%; Score 10; DB 20; Length 400;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CCTATGAG 10  
 |||||||||  
 Db 315 CCTATGAG 324  
 RESULT 13  
 LOCUS 104785 413 bp mRNA EST 06-NOV-1997  
 DEFINITION 833 lambda-PRL2 Arabidopsis thaliana cDNA clone 3d177p, mRNA sequence.  
 ACCESSION 104785  
 NID 9315945  
 VERSION 104785.1 GI:315945  
 KEYWORDS EST.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;

Arbidoopsis.  
 1 (bases 1 to 413)  
 REFERENCE Newman,T., deBrujn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M., Retzel,E. and Somerville,C.  
 TITLE Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones Plant Physiol. 106, 1241-1255 (1994)  
 JOURNAL MEDLINE 95148729  
 COMMENT  
 CONTACT: Thomas Newman  
 MSU-DOE Plant Research Laboratory  
 Michigan State University  
 MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E. Lansing, MI  
 Tel: 517-353-0854  
 Fax: 517-353-9168  
 Email: 22313cne@dm.cl.msu.edu  
 Seq primer: T7 dye primer.  
 Location/Qualifiers  
 1..413  
 /organism="Arabidopsis thaliana"  
 /strain="var columbia"  
 /db\_xref="taxon:3702"  
 /clone="3d177p"  
 /clone\_1db="lambda-PRL2"  
 /note="Vector: lambda Zip-lox; Site\_1: Sal; Site\_2: Not; Lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques). The vector is BRV's lambda Zip-lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dt primed cDNA."  
 BASE COUNT 100 a 91 c 100 g 117 t 5 others  
 ORIGIN  
 Query Match 100.0%; Score 10; DB 20; Length 413;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CCTATGAG 10  
 |||||||||  
 Db 79 CCTATGAG 88  
 RESULT 14  
 LOCUS 106534 357 bp mRNA EST 30-JUN-1993  
 DEFINITION EST04423 Fetal brain, Strata gene (cat#936206) Homo sapiens cDNA clone HFB078 similar to lymphocyte function associated antigen-3, mRNA sequence.  
 ACCESSION 106534  
 NID 9317683  
 VERSION 106534.1 GI:317683  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE Adams,M.D., Keriavage,A.R., Fields,C. and Venter,J.C.  
 TITLE 1 (bases 1 to 357)  
 JOURNAL 3,400 expressed sequence tags identify diversity of transcripts from human brain  
 MEDLINE Nature Genet. 4, 256-267 (1993)  
 COMMENT  
 Contact: Adams, MD  
 The Institute for Genomic Research  
 932 Clopper Road, Gaithersburg, MD 20878

Tel: 3018699056  
Fax: 3018699423

Email: mdadams@tigr.org  
Seq primer: M13-21.

FEATURES  
Source Location/Qualifiers

1..357  
/organism="Homo sapiens"  
/db\_xref="ATCC (Inhost):83193"  
/db\_xref="taxon:9606"  
/clone="HFDU78"  
/note="Vector: Fetal brain, Stratagene (cat#936206)"  
/note="Vector: LambdaZAP-II, 17-18 wk gestation, female;  
oligo-dt + random primed cDNA synthesis; lambdaZAP-II  
vector, 1.0kb average insert size."  
BASE COUNT 114 a 67 c 68 g 106 t 2 others  
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 357;  
Best Local Similarity 100.0%; Pred. No. 2e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTATGGAG 10  
|||||  
DB 341 CCTATGGAG 350

RESULT 15  
T13875/c

LOCUS T13875 453 bp mRNA EST 07-JAN-1998  
DEFINITION 2040 Lambda-PRL2 Arabidopsis thaliana cDNA clone 426217, mRNA  
sequence.

ACCESSION T13875  
T13875  
9430244

VERSION T13875.1 GI:430244

KEYWORDS EST

SOURCE

ORGANISM thale cress,  
Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
Arabidopsis.

REFERENCE 1 (bases 1 to 453)

AUTHORS

Newman,T., deBry,J.N., F.J., Green,P., Keegstra,K., Kende,H.,  
McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M.,  
Reisel,E. and Somerville,C.  
Genes galore: a summary of methods for accessing results from  
large-scale partial sequencing of anonymous Arabidopsis cDNA clones  
Plant Physiol. 106, 1241-1255 (1994)  
95148729

COMMENT

Contact: Thomas Newman  
MSU-DOE Plant Research Laboratory  
Michigan State University

MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.  
Lansing, MI

Tel: 517-353-0854

Fax: 517-353-9168

Email: 22313cne@lhm.cl.msu.edu

Seq primer: T7.

FEATURES  
Source

Location/Qualifiers  
1..453  
/organism="Arabidopsis thaliana"  
/strain="var Columbia"  
/db\_xref="taxon:3702"  
/clone="426217"  
/clone.lib="lambda-PRL2"  
/note="Vector: lambda ZAP-lox; Site\_1: Sal; Site\_2: Not;  
lambda PRL2 is a cDNA library derived from equal  
quantities of 4 pools of mRNA. The mRNA sources were 1) 7  
day germinated etiolated seedlings; 2) tissue culture  
grown roots; 3) staged plants half with 24 hour light  
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)

same plants as 3 but aerial tissue (stems, flowers and  
siliques. The vector is BRU's lambda ZAP-lox. The cDNA  
inserts were directionally cloned with Sal-Not arms using  
oligo dt primed cDNA.  
BASE COUNT 115 a 83 c 115 g 130 t 10 others  
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 453;  
Best Local Similarity 100.0%; Pred. No. 2.1e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTATGGAG 10  
|||||  
DB 205 CCTATGGAG 196

Search completed: September 17, 1999, 21:28:03  
Job time: 14267 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:01:35 ; Search time 1962.17 Seconds  
(without alignments)  
16.208 Million cell updates/sec

Title: US-08-956-518a-91

Perfect score: 10

Sequence: 1 TATGTTTAG 10

Scoring table: IDENTITY\_NUC

Indexed: 679419 segs, 1590154680 residues

Database :

GenEmbl:.\*  
1: gb\_dal:.\*  
2: gb\_daz:.\*  
3: gb\_om:.\*  
4: gb\_ov:.\*  
5: gb\_pac:.\*  
6: gb\_ph:.\*  
7: gb\_p11:.\*  
8: gb\_p12:.\*  
9: gb\_p13:.\*  
10: gb\_p14:.\*  
11: gb\_p15:.\*  
12: gb\_p16:.\*  
13: gb\_p17:.\*  
14: gb\_p18:.\*  
15: gb\_p19:.\*  
16: gb\_p20:.\*  
17: gb\_p21:.\*  
18: gb\_p22:.\*  
19: gb\_p23:.\*  
20: gb\_p24:.\*  
21: gb\_p25:.\*  
22: gb\_p26:.\*  
23: gb\_p27:.\*  
24: gb\_p28:.\*  
25: gb\_p29:.\*  
26: gb\_p30:.\*  
27: gb\_p31:.\*  
28: gb\_p32:.\*  
29: gb\_p33:.\*  
30: gb\_p34:.\*  
31: gb\_p35:.\*  
32: gb\_p36:.\*  
33: gb\_p37:.\*  
34: gb\_p38:.\*  
35: gb\_p39:.\*  
36: gb\_p40:.\*  
37: gb\_p41:.\*  
38: gb\_p42:.\*  
39: gb\_p43:.\*  
40: gb\_p44:.\*  
41: gb\_p45:.\*  
42: gb\_p46:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	100.0	1709	1 AAPSBADNA	X64174 A.azollae p

c	2	10	100.0	1507	1	AB001778	AB001778 Chlamydia
c	3	10	100.0	1507	1	AB001779	AB001779 Chlamydia
c	4	10	100.0	1507	1	AB001780	AB001780 Chlamydia
c	5	10	100.0	1507	1	AB001781	AB001781 Chlamydia
c	6	10	100.0	1507	1	AB001782	AB001782 Chlamydia
c	7	10	100.0	1507	1	AB001783	AB001783 Chlamydia
c	8	10	100.0	1507	1	AB001784	AB001784 Chlamydia
c	9	10	100.0	1507	1	AB001785	AB001785 Chlamydia
c	10	10	100.0	1507	1	AB001786	AB001786 Chlamydia
c	11	10	100.0	1507	1	AB001787	AB001787 Chlamydia
c	12	10	100.0	1507	1	AB001788	AB001788 Chlamydia
c	13	10	100.0	1507	1	AB001789	AB001789 Chlamydia
c	14	10	100.0	1507	1	AB001790	AB001790 Chlamydia
c	15	10	100.0	1507	1	AB001791	AB001791 Chlamydia
c	16	10	100.0	1507	1	AB001802	AB001802 Chlamydia
c	17	10	100.0	1507	1	AB001809	AB001809 Chlamydia
c	18	10	100.0	1507	1	AB001810	AB001810 Chlamydia
c	19	10	100.0	1507	1	AB001811	AB001811 Chlamydia
c	20	10	100.0	1507	1	AB001812	AB001812 Chlamydia
c	21	10	100.0	1507	1	AB001813	AB001813 Chlamydia
c	22	10	100.0	1440	1	AB009228	AB009228 Chlamydia
c	23	10	100.0	585	1	AB011313	AB011313 Chlamydia
c	24	10	100.0	4867	1	AB019247	AB019247 Chlamydia
c	25	10	100.0	4543	1	AB019249	AB019249 Chlamydia
c	26	10	100.0	4247	1	ACNAPPAB	ACNAPPAB Chlamydia
c	27	10	100.0	5085	1	ACPOOSYN	ACPOOSYN Chlamydia
c	28	10	100.0	3165	1	ALTNAG096	ALTNAG096 Chlamydia
c	29	10	100.0	287000	1	AP000001	AP000001 Pyrococcus
c	30	10	100.0	217000	1	AP000004	AP000004 Pyrococcus
c	31	10	100.0	5120	1	APCLYI	APCLYI Chlamydia
c	32	10	100.0	8292	1	APHLXTRAD	APHLXTRAD Chlamydia
c	33	10	100.0	3189	1	ASU14553	ASU14553 Chlamydia
c	34	10	100.0	1895	1	BAC17LPEN	BAC17LPEN Chlamydia
c	35	10	100.0	2878	1	BACBCT	BACBCT Chlamydia
c	36	10	100.0	1892	1	BACDEP	BACDEP Chlamydia
c	37	10	100.0	2041	1	BACHEMIBC	BACHEMIBC Chlamydia
c	38	10	100.0	282700	1	BACJH642	BACJH642 Chlamydia
c	39	10	100.0	1841	1	BACMSB	BACMSB Chlamydia
c	40	10	100.0	1324	1	BACMSBRS	BACMSBRS Chlamydia
c	41	10	100.0	1495	1	BACMSERND	BACMSERND Chlamydia
c	42	10	100.0	23286	1	BACPONAYPP	BACPONAYPP Chlamydia
c	43	10	100.0	48032	1	BACSKIN	BACSKIN Chlamydia
c	44	10	100.0	3479	1	BACTOXB	BACTOXB Chlamydia
c	45	10	100.0	193772	42	AC007073	AC007073 Homo sapi

## ALIGNMENTS

RESULT 1  
LOCUS AAPSBADNA/c  
DEFINITION A.azollae psbA gene for D1 protein.  
ACCESSION X64174  
VERSION 938650  
KEYWORDS D1 protein; P680-apoprotein; photosystem II 32 kD protein; Photosystem II reaction centre; psbA gene.  
SOURCE  
ORGANISM Anabaena azollae  
REFERENCE  
1 (bases 1 to 1709)  
AUTHORS Nierzwicki-Bauer, S.A.  
TITLE Direct Submission  
JOURNAL Submitted (15-JAN-1992) S.A. Nierzwicki-Bauer, Rensselaer Polytechnic Institute, Plant Science Group, Biology Department, 110 8th Street, Troy, New York, 12180, USA  
REFERENCE  
2 (bases 1 to 1709)  
AUTHORS Gebhardt, J.S. and Nierzwicki-Bauer, S.A.  
TITLE Characterization of psbA gene family expression in symbiotic Anabaena azollae  
JOURNAL Unpublished  
FEATURES  
Location/Qualifiers



```

REFERENCE      2 (bases 1 to 1507)
AUTHORS        Takahashi,T., Masuda,M., Tsuruno,T., Mori,Y., Takashima,I. and
                Kikuchi,N.
TITLE          Phylogenetic analyses of Chlamydia psittaci from birds based on the
                16S rRNA sequence
JOURNAL        Unpublished (1997)
FEATURES       location/Qualifiers
                1..1507
                /organism="Chlamydia psittaci"
                /isolate="Bud-11F from a budgerigar"
                /db_xref="taxon:812"
                rRNA
                1..1507
                /product="16S rRNA"
BASE COUNT     405 a 312 c 441 g 349 t
ORIGIN
Query Match    100.0%; Score 10; DB 1; Length 1507;
                Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10
    |||||
Db 193 TATGTTTAG 184

RESULT 5
AB001781 1507 bp DNA BCT 06-MAR-1998
LOCUS     Chlamydia psittaci gene for 16S rRNA, isolate Bud-16F.
DEFINITION
ACCESSION AB001781
VERSION   91902844
KEYWORDS  16S ribosomal RNA.
SOURCE    Chlamydia psittaci (isolate:Bud-16F from a budgerigar) DNA.
ORGANISM  Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE  1 (bases 1 to 1507)
AUTHORS    Takahashi,T.
TITLE      Direct Submission
JOURNAL    Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
          Tatsufumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen
          Univ., Epizootiology; 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido
          069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)
REFERENCE  2 (bases 1 to 1507)
AUTHORS    Takahashi,T., Masuda,M., Tsuruno,T., Mori,Y., Takashima,I. and
          Kikuchi,N.
TITLE      Phylogenetic analyses of Chlamydia psittaci from birds based on the
          16S rRNA sequence
JOURNAL    Unpublished (1997)
FEATURES   location/Qualifiers
            1..1507
            /organism="Chlamydia psittaci"
            /isolate="Bud-16F from a budgerigar"
            /db_xref="taxon:812"
            rRNA
            1..1507
            /product="16S rRNA"
BASE COUNT 405 a 312 c 441 g 349 t
ORIGIN

```

```

DEFINITION Chlamydia psittaci gene for 16S rRNA, isolate:Bud-5695.
ACCESSION  AB001782
VERSION    91902845
KEYWORDS   16S ribosomal RNA.
SOURCE     Chlamydia psittaci (isolate:Bud-5695 from a budgerigar) DNA.
ORGANISM   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE  1 (bases 1 to 1507)
AUTHORS    Takahashi,T.
TITLE      Direct Submission
JOURNAL    Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
          Tatsufumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen
          Univ., Epizootiology; 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido
          069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)
REFERENCE  2 (bases 1 to 1507)
AUTHORS    Takahashi,T., Masuda,M., Tsuruno,T., Mori,Y., Takashima,I. and
          Kikuchi,N.
TITLE      Phylogenetic analyses of Chlamydia psittaci from birds based on the
          16S rRNA sequence
JOURNAL    Unpublished (1997)
FEATURES   location/Qualifiers
            1..1507
            /organism="Chlamydia psittaci"
            /isolate="Bud-5695 from a budgerigar"
            /db_xref="taxon:812"
            rRNA
            1..1507
            /product="16S rRNA"
BASE COUNT 405 a 312 c 441 g 349 t
ORIGIN
Query Match    100.0%; Score 10; DB 1; Length 1507;
                Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10
    |||||
Db 193 TATGTTTAG 184

RESULT 7
AB001783 1507 bp DNA BCT 14-MAY-1999
LOCUS     Chlamydia abortus gene for 16S rRNA, strain B577.
DEFINITION
ACCESSION AB001783
VERSION   91902846
KEYWORDS  16S ribosomal RNA.
SOURCE    Chlamydia abortus (strain:B577, ATCC VR-656, specific:host:Ovis
          aries) DNA.
ORGANISM   Chlamydia abortus
REFERENCE  1 (bases 1 to 1507)
AUTHORS    Takahashi,T., Masuda,M., Tsuruno,T., Mori,Y., Takashima,I. and
          Kikuchi,N.
TITLE      Phylogenetic analyses of Chlamydia psittaci from birds based on the
          16S rRNA sequence
JOURNAL    Unpublished (1997)
REFERENCE  2 (bases 1 to 1507)
AUTHORS    Takahashi,T.
TITLE      Direct Submission
JOURNAL    Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
          Tatsufumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen
          Univ., Epizootiology; 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido
          069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)
FEATURES   location/Qualifiers
            1..1507
            /organism="Chlamydia abortus"
            /specific_host="Ovis aries"
            /strain="B577, ATCC VR-656"
            /db_xref="taxon:83555"
            /note="submitted as Chlamydia psittaci"

```



RNA 1..1507  
 /product="16S rRNA"  
 BASE COUNT 408 a 312 c 438 g 349 t  
 ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 1507;  
 Best Local Similarity 100.0%; Pred. No. 1e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10  
 |||||  
 DB 193 TATGTTTAG 184

RESULT 8  
 AB001786/c 1507 bp DNA BCT 06-MAR-1998  
 LOCUS Chlamydia psittaci gene for 16S rRNA, strain:GCP-1.  
 DEFINITION  
 VERSION 91902849

KEYWORDS AB001786.1 GI:1902849  
 16S ribosomal RNA.  
 SOURCE Chlamydia psittaci (strain:GCP-1) DNA.  
 ORGANISM Chlamydia psittaci

REFERENCE 1 (bases 1 to 1507)  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia:  
 Takahashi, T.  
 AUTHORS Direct Submission  
 JOURNAL Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.  
 Tatsufumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen  
 Univ., Epizootiology, 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido  
 069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)

REFERENCE 2 (bases 1 to 1507)  
 Takahashi, T., Masuda, M., Tsuruno, T., Mori, Y., Takashima, I. and  
 Kikuchi, N.  
 TITLE Phylogenetic analyses of Chlamydia psittaci from birds based on the  
 16S rDNA sequence  
 JOURNAL Unpublished (1997)  
 FEATURES Location/Qualifiers

source 1..1507  
 /organism="Chlamydia psittaci"  
 /strain="GCP-1"  
 /db\_xref="taxon:812"

RNA 1..1507  
 /product="16S rRNA"  
 BASE COUNT 405 a 312 c 441 g 349 t  
 ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 1507;  
 Best Local Similarity 100.0%; Pred. No. 1e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10  
 |||||  
 DB 193 TATGTTTAG 184

RESULT 9  
 AB001787/c 1507 bp DNA BCT 06-MAR-1998  
 LOCUS Chlamydia psittaci gene for 16S rRNA, strain:Itch.  
 DEFINITION  
 ACCESSION AB001787  
 NID 91902850

VERSION AB001787.1 GI:1902850  
 16S ribosomal RNA.  
 KEYWORDS Chlamydia psittaci (strain:Itch) DNA.  
 SOURCE Chlamydia psittaci

ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia:  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia:  
 Takahashi, T.  
 AUTHORS Direct Submission  
 JOURNAL

## JOURNAL

Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.  
 Tatsufumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen  
 Univ., Epizootiology, 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido  
 069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)

REFERENCE 2 (bases 1 to 1507)  
 Takahashi, T., Masuda, M., Tsuruno, T., Mori, Y., Takashima, I. and  
 Kikuchi, N.  
 TITLE Phylogenetic analyses of Chlamydia psittaci from birds based on the  
 16S rDNA sequence  
 JOURNAL Unpublished (1997)  
 FEATURES Location/Qualifiers

source 1..1507  
 /organism="Chlamydia psittaci"  
 /strain="Itch"  
 /db\_xref="taxon:812"

RNA 1..1507  
 /product="16S rRNA"  
 BASE COUNT 405 a 312 c 441 g 349 t  
 ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 1507;  
 Best Local Similarity 100.0%; Pred. No. 1e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10  
 |||||  
 DB 193 TATGTTTAG 184

RESULT 10  
 AB001788/c 1507 bp DNA BCT 06-MAR-1998  
 LOCUS Chlamydia psittaci gene for 16S rRNA, isolate:Izawa-1.  
 DEFINITION  
 ACCESSION AB001788

NID 91902851  
 VERSION AB001788.1 GI:1902851  
 16S ribosomal RNA.  
 KEYWORDS Chlamydia psittaci (isolate:Izawa-1 from a budgerigar) DNA.  
 SOURCE Chlamydia psittaci

ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia:  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia:  
 Takahashi, T.  
 AUTHORS Direct Submission  
 JOURNAL Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.  
 Tatsufumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen  
 Univ., Epizootiology, 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido  
 069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)

REFERENCE 2 (bases 1 to 1507)  
 Takahashi, T., Masuda, M., Tsuruno, T., Mori, Y., Takashima, I. and  
 Kikuchi, N.  
 TITLE Phylogenetic analyses of Chlamydia psittaci from birds based on the  
 16S rDNA sequence  
 JOURNAL Unpublished (1997)  
 FEATURES Location/Qualifiers

source 1..1507  
 /organism="Chlamydia psittaci"  
 /isolate="Izawa-1 from a budgerigar"  
 /db\_xref="taxon:812"

RNA 1..1507  
 /product="16S rRNA"  
 BASE COUNT 405 a 312 c 441 g 349 t  
 ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 1507;  
 Best Local Similarity 100.0%; Pred. No. 1e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10  
 |||||  
 DB 193 TATGTTTAG 184

```

RESULT 11
LOCUS AB001789/c 1507 bp DNA BCT 06-MAR-1998
DEFINITION Chlamydia psittaci gene for 16S rRNA, strain:Koala.
ACCESSION AB001789
NID g1902852
VERSION AB001789.1 GI:1902852
KEYWORDS 16S ribosomal RNA.
SOURCE Chlamydia psittaci (strain:Koala) DNA.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
AUTHORS Takahashi,T.
TITLE 1 (bases 1 to 1507)
JOURNAL Direct Submission
FEATURES
SOURCE Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
Tatsufumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen
Univ., Epizootiology; 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido
069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)
2 (bases 1 to 1507)
Takahashi,T., Masuda,M., Tsuruno,T., Mori,Y., Takashima,I. and
Kikuchi,N.
Phylogenetic analyses of Chlamydia psittaci from birds based on the
16S rDNA sequence
Unpublished (1997)
LOCATION/Qualifiers
1..1507
/organism="Chlamydia psittaci"
/strain="Koala"
/db_xref="taxon:812"
/product="16S rRNA"
RNA 1..1507
BASE COUNT 405 a 312 c 441 g 349 t
ORIGIN
Query Match 100.0%; Score 10; DB 1; Length 1507;
Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TATGTTTAG 10
Db 193 TATGTTTAG 184
---
RESULT 12
LOCUS 01790/c 1507 bp DNA BCT 06-MAR-1998
DEFINITION Chlamydia psittaci gene for 16S rRNA, isolate:Mizuno-1F.
ACCESSION AB001790
NID g1902853
VERSION AB001790.1 GI:1902853
KEYWORDS 16S ribosomal RNA.
SOURCE Chlamydia psittaci (isolate:Mizuno-1F from a budgerigar) DNA.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
AUTHORS Takahashi,T.
TITLE 1 (bases 1 to 1507)
JOURNAL Direct Submission
FEATURES
SOURCE Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
Tatsufumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen
Univ., Epizootiology; 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido
069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)
2 (bases 1 to 1507)
Takahashi,T., Masuda,M., Tsuruno,T., Mori,Y., Takashima,I. and
Kikuchi,N.
Phylogenetic analyses of Chlamydia psittaci from birds based on the
16S rDNA sequence
Unpublished (1997)
LOCATION/Qualifiers
1..1507
/organism="Chlamydia psittaci"
/strain="Om1ya"
/db_xref="taxon:812"
/product="16S rRNA"
RNA 1..1507
BASE COUNT 404 a 312 c 442 g 349 t
ORIGIN
Query Match 100.0%; Score 10; DB 1; Length 1507;
Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TATGTTTAG 10
Db 193 TATGTTTAG 184
---
RESULT 13
LOCUS AB001791/c 1507 bp DNA BCT 06-MAR-1998
DEFINITION Chlamydia psittaci gene for 16S rRNA, strain:Om1ya.
ACCESSION AB001791
NID g1902854
VERSION AB001791.1 GI:1902854
KEYWORDS 16S ribosomal RNA.
SOURCE Chlamydia psittaci (strain:Om1ya) DNA.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
AUTHORS Takahashi,T.
TITLE 1 (bases 1 to 1507)
JOURNAL Direct Submission
FEATURES
SOURCE Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
Tatsufumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen
Univ., Epizootiology; 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido
069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)
2 (bases 1 to 1507)
Takahashi,T., Masuda,M., Tsuruno,T., Mori,Y., Takashima,I. and
Kikuchi,N.
Phylogenetic analyses of Chlamydia psittaci from birds based on the
16S rDNA sequence
Unpublished (1997)
LOCATION/Qualifiers
1..1507
/organism="Chlamydia psittaci"
/strain="Om1ya"
/db_xref="taxon:812"
/product="16S rRNA"
RNA 1..1507
BASE COUNT 404 a 312 c 442 g 349 t
ORIGIN
Query Match 100.0%; Score 10; DB 1; Length 1507;
Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TATGTTTAG 10
Db 193 TATGTTTAG 184
---
RESULT 14
LOCUS AB001802/c 1507 bp DNA BCT 06-MAR-1998
DEFINITION Chlamydia psittaci gene for 16S rRNA, strain:PgAu46.
ACCESSION AB001802
NID g1902865
VERSION AB001802.1 GI:1902865
KEYWORDS 16S ribosomal RNA.
SOURCE Chlamydia psittaci (strain:PgAu46) DNA.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
AUTHORS Takahashi,T.
TITLE 1 (bases 1 to 1507)
JOURNAL Direct Submission
FEATURES
SOURCE Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
Tatsufumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen
Univ., Epizootiology; 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido
069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)
2 (bases 1 to 1507)
Takahashi,T., Masuda,M., Tsuruno,T., Mori,Y., Takashima,I. and
Kikuchi,N.
Phylogenetic analyses of Chlamydia psittaci from birds based on the
16S rDNA sequence
Unpublished (1997)
LOCATION/Qualifiers
1..1507
/organism="Chlamydia psittaci"
/strain="PgAu46"
/db_xref="taxon:812"
/product="16S rRNA"
RNA 1..1507
BASE COUNT 404 a 312 c 442 g 349 t
ORIGIN
Query Match 100.0%; Score 10; DB 1; Length 1507;
Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TATGTTTAG 10
Db 193 TATGTTTAG 184

```

TITLE Direct Submission  
JOURNAL Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.

REFERENCE Tatsuimi Takahashi, School of Veterinary Medicine, Rakuno Gakuen Univ., Epizootiology, 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido 069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)

AUTHORS 2 (bases 1 to 1507)  
Takahashi, T., Masuda, M., Tsuruno, T., Mori, Y., Takashima, I. and Kikuchi, N.

TITLE Phylogenetic analyses of Chlamydia psittaci from birds based on the 16S rDNA sequence  
JOURNAL Unpublished (1997)

FEATURES Location/Qualifiers  
source 1..1507

/organism="Chlamydia psittaci"  
/strain="PgAu46"  
/db\_xref="taxon:812"

RNA 1..1507

BASE COUNT 403 a 311 c 443 g 350 t

ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 1507;

Best Local Similarity 100.0%; Pred. No. 1e+04; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10

DB 193 TATGTTTAG 184

RESULT 15  
AB001809 1507 bp DNA BCT 06-MAR-1998

LOCUS Chlamydia psittaci gene for 16S rRNA, strain:PrK46.

DEFINITION AB001809

ACCESSION 91902872

VERSION AB001809.1 GI:1902872

KEYWORDS 16S ribosomal RNA.

SOURCE Chlamydia psittaci (strain:PrK46) DNA.

ORGANISM Chlamydia psittaci

REFERENCE 1 (bases 1 to 1507)

AUTHORS Takahashi, T.

TITLE Direct Submission

JOURNAL Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.

REFERENCE Tatsuimi Takahashi, School of Veterinary Medicine, Rakuno Gakuen Univ., Epizootiology, 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido 069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)

AUTHORS 2 (bases 1 to 1507)  
Takahashi, T., Masuda, M., Tsuruno, T., Mori, Y., Takashima, I. and Kikuchi, N.

TITLE Phylogenetic analyses of Chlamydia psittaci from birds based on the 16S rDNA sequence

JOURNAL Unpublished (1997)

FEATURES Location/Qualifiers  
source 1..1507

/organism="Chlamydia psittaci"  
/strain="PrK46"  
/db\_xref="taxon:812"

RNA 1..1507

BASE COUNT 407 a 313 c 439 g 348 t

ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 1507;

Best Local Similarity 100.0%; Pred. No. 1e+04; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10

DB 193 TATGTTTAG 184

Search completed: September 17, 1999, 22:01:36  
Job time: 16279 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:33:35 ; Search time 425.19 Seconds

(Without alignments)  
5.884 Million cell updates/sec

Title: US-08-956-518A-91

Perfect score: 10

Sequence: 1 TATGTTTAG 10

Scoring table: IDENTITY\_NUC

arched: 311585 segs, 125096042 residues

-database : N.Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	100.0	9360	1 Q02830	CDNA to HIV-2 RNA.
2	10	100.0	3060	1 Q03098	Enod2a genomic clo
3	10	100.0	3856	1 Q03097	Enod2b genomic clo
4	10	100.0	18177	1 N90490	DNA of human retin
5	10	100.0	653	1 N90728	Total CDNA from cl
6	10	100.0	9822	1 N91774	Entire HIV-2/ST pr
7	10	100.0	774	1 N91602	Partial nucleotide
8	10	100.0	1975	1 Q05057	Drai/Drai fragment
9	10	100.0	1355	1 N82025	Fragment containin
10	10	100.0	9643	1 N80859	Sequence of entire
11	10	100.0	9236	1 N80436	Entire sequence of
12	10	100.0	1566	1 N82198	GAGRON sequence f
13	10	100.0	1566	1 N70887	HIV virus gag sequ
14	10	100.0	4355	1 Q10947	Bovine Coronavirus
15	10	100.0	35	1 Q10093	Probe 782 to Chlam
16	10	100.0	4462	1 Q11657	3' Terminal of TSM
17	10	100.0	9192	1 Q11943	Nucleotide sequenc
18	10	100.0	40	1 Q12325	HPV type 33-specif
19	10	100.0	978	1 Q13847	Phibacin PSBX muta
20	10	100.0	983	1 Q13846	Phibacin PSBX orfl
21	10	100.0	1200	1 Q13848	Phibacin PSBX 1.2k
22	10	100.0	1201	1 Q13849	Phibacin PSBX 1.2k
23	10	100.0	3458	1 Q14256	51.4 and 41.9 kD t
24	10	100.0	2411	1 Q14658	bt1109P gene. Bac1
25	10	100.0	15155	1 N50107	DNA encoding Facto
26	10	100.0	14704	1 Q20685	PKS 741 insert con
27	10	100.0	14704	1 Q20685	PKS 741 insert con
28	10	100.0	8937	1 Q20602	Net gene. DNA sequ
29	10	100.0	9672	1 Q20616	ROD HIV-2 isolate
30	10	100.0	9709	1 Q22488	HIV-1 proviral clo
31	10	100.0	7432	1 Q22965	True type 3 poliov
32	10	100.0	25	1 Q22047	Sequence of PCR pr
33	10	100.0	3209	1 Q22769	Canarypox PvuII fr
34	10	100.0	2949	1 Q25764	Canarypox virus C5
35	10	100.0	1105	1 Q23894	Encodes human prot
36	10	100.0	7323	1 Q25033	Bovine RSV strain
37	10	100.0	1260	1 Q22042	Sequence encoding
38	10	100.0	639	1 Q26962	Human PAP 3'-UTR.
39	10	100.0	5008	1 Q27223	MSF precursor. New
40	10	100.0	3209	1 Q29107	Canarypox PvuII fr
41	10	100.0	1521	1 Q29107	Sequence of the fl
42	10	100.0	7431	1 Q50148	Attenuated (4720,
43	10	100.0	7399	1 Q33479	HPBACR1. Defective

## ALIGNMENTS

RESULT 1	002830/c	ID	002830 standard; cdna; 9360 BP.
AC	002830/	DT	29-MAY-1989 (first entry)
DE	CDNA to HIV-2 RNA.	KW	HIV, AIDS; Vaccine; PUC-HIV-2(GH-1).
OS	HIV-2.	FS	Key
FT	cds	FT	Location/Qualifiers
FT	cds	FT	544..2109
FT	cds	FT	/*tag- a
FT	cds	FT	/label-Gag reading frame
FT	cds	FT	1830..4934
FT	cds	FT	/*tag- b
FT	cds	FT	/label-Pol gene reading frame.
FT	cds	FT	4867..5511
FT	cds	FT	/*tag- c
FT	cds	FT	/label-Q gene reading frame
FT	cds	FT	5342..5677
FT	cds	FT	/*tag- d
FT	cds	FT	/label-X gene reading frame
FT	cds	FT	5680..5994
FT	cds	FT	/*tag- e
FT	cds	FT	/label-R gene reading frame
FT	cds	FT	6148..8703
FT	cds	FT	/*tag- f
FT	cds	FT	/label-Env gene reading frame.
FT	cds	FT	8540..9304
FT	cds	FT	/*tag- g
FT	cds	FT	/label-F gene reading frame.
PN	J01289486-A.	PD	21-NOV-1989.
PR	16-MAY-1988; 119024.	PR	16-MAY-1988; JP-119024.
PA	Toa Nentryo Kogyo KK, Fuji Kabi KK.	DR	WPJ; 90-005177/01.
DR	P-PSDB; R04024-30.	PT	DNA indicating complement to RNA gene -
PT	of Human Immunodeficiency Virus type 2 used for new vaccine or	PT	diagnostic for AIDS virus.
PS	Claim 2; Fig.4; 12pp; Japanese.	CC	CDNA to novel HIV-2 (GH-1) has been integrated into plasmid
CC	PUC HIV-2(GH-1). Useful for diagnosis and vaccination against the virus.	CC	Carries 7 overlapping genes in various reading frames, including gag,
CC	Pol and Env.	SQ	Sequence 9360 BP; 3211 A; 1944 C; 2317 G; 1888 T;
QY	1 TATGTTTAG 10	DB	643 TATGTTTAG 634
DB	643 TATGTTTAG 634		
RESULT 2	003098/c	ID	003098 standard; cdna; 3060 BP.
AC	003098/	DT	05-JUN-1990 (first entry)
DE	Enod2a genomic clone.	KW	Enod2; soybean; root nodule; regulatory region; nodulin 75; ss.
OS	Glycine max (L) Merr. cv. Williams.	FS	Key
FT	cds	FT	Location/Qualifiers



```

FT      exon 20      11809..11955
FT      intron (4.0 kb)
FT      exon 21      12278..12279
FT      intron (0.2 kb) 12538..12642
FT      exon 22      12948/12949
FT      exon 23      13204..13317
FT      intron (7.4 kb) 13411..13574
FT      exon 24      13883/13884
FT      intron (2.8 kb) 14209..14239
FT      exon 25      14641/14642
FT      exon 26      14834..14977
FT      intron (1.7 kb) 15488..15538
FT      exon 27      15952/15953
FT      polyA signal 16143..16213
FT      18010

PN      M08906703-A.
PD      27-JUL-1989.
PF      23-JAN-1989; U00293.
PI      21-JAN-1988; US-146525.
PR      (DRYJ) Dryja T P.
PS      WPI: 89-233856/32.
PT      Dryja T P; Friend S; Vandell D W.
DR      P-PSDB: P90599.
PT      New human retino-blastoma gene and polypeptide(s)
PS      - used for screening individuals for defective retino-blastoma
PT      gene and treating these patients.
PS      Disclosure, fig 6; 71 pp. English.
CC      DNA encoding human retinoblastoma (RB) gene exons and flanking regions.
CC      The cDNA (see P90599 and N90490) can be used to screen individuals
CC      for the presence of the mutated RB gene. The RB polypeptide
CC      can prevent retinoblastoma formation, and corresp. antibodies
CC      can be used in tumour immunodiagnosis. Refer to patent for more
CC      sequence analysis details.
SQ      Sequence 18177 BP; 5765 A; 2997 C; 3172 G; 6202 T; 41 Others.

Query Match      100.0%; Score 10; DB 1; Length 18177;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TATGTTTAG 10
DB      15207 TATGTTTAG 15216

RESULT 5
N90728 standard; DNA; 653 BP.
N90728;
DE      16-JUN-1990 (first entry)
DE      Total cDNA from clone for the MIP-1 beta component of murine
DE      inflammatory cytokine
DE      Inflammatory cytokine; MIP-1 beta component;
KW      mouse macrophage cell line RAW 264.7.
OS      Mouse
FH      Key      Location/Qualifiers
FT      cds      63..341
FT      mat_peptide 132..341
FT      /*tag- a
FT      /*tag- b
PN      EP-310136-A.
PD      05-APR-1989.
PF      03-OCT-1988; 116329
PR      02-SEP-1988; US-238937; US-104827.
PA      (UYRO-) Rockefeller Univ.
PI      Cerami A, Beutler B, Wolpe SD;
DR      WPI: 89-101457/14.
DR      P-PSDB: P93591.
PT      New cytokine which induces inflammation -
PT      obd. by purification of mediator substance obd. in supernatant of
PT      stimulated macrophage cells
PS      Figure 15; 46pp; English.
CC      Cytokine MIP-1 is composed of two components, MIP-1alpha and MIP-1beta.
CC      The MIP-1 alpha was initially recovered from murine cells (mouse

```

```

CC      macrophage cell line RAW 264.7). Inflammatory cytokine may be used to
CC      identify idiopathic or invasive stimuli or determining the presence of
CC      stimulated, spontaneous or idiopathic pathological states in mammals. It
CC      induces fever in rabbits and superoxide formation in vitro in human
CC      neutrophils.
SQ      Sequence 653 BP; 139 A; 174 C; 141 G; 199 T;

Query Match      100.0%; Score 10; DB 1; Length 653;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TATGTTTAG 10
DB      520 TATGTTTAG 529

RESULT 6
N91774/c
ID      N91774 standard; DNA; 9822 BP.
AC      N91774;
DE      16-MAR-1990 (first entry)
DE      Entire HIV-2/ST provirus DNA
KW      Antibodies
OS      Human immunodeficiency virus-2.
PN      W08909815-A.
PD      19-OCT-1989.
PR      04-APR-1989; U001379.
PS      04-APR-1988; US-177567.
PA      (RESF) Research Corporation Technologies Inc.
PI      Shaw GM, Hahn BH, Kong L, Lee SW;
DR      WPI: 89-324223/44.
PT      Novel virus of the HIV-2 family + used in the prodn. of diagnostic
PT      reagents, vaccines and therapies for AIDS and AIDS-related diseases
PS      Claim 37; page 47-53; 66pp; English.
CC      This is used as a source of recombinant viral components in the
CC      development of diagnostic assays for HIV-2 viruses.
SQ      Sequence 9822 BP; 3359 A; 2023 C; 2441 G; 1999 T;

Query Match      100.0%; Score 10; DB 1; Length 9822;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TATGTTTAG 10
DB      647 TATGTTTAG 638

RESULT 7
N91602
ID      N91602 standard; DNA; 774 BP.
AC      N91602;
DE      17-JUL-1990 (first entry)
DE      Partial nucleotide sequence (5' end) of human papilloma virus (HPV)
DE      type 33 (HPV-33)
KW      Human papilloma virus; type 33; in situ hybridisation assay;
KW      cellular smear; cervical carcinoma.
OS      Human papilloma virus.
PN      W08902934-A.
PD      06-APR-1989.
PF      30-SEP-1988; U03367.
PR      02-OCT-1987; US-103979.
PA      (MICR-) Microprobe Corp.
PI      Schwartz DE, Adams TH;
DR      WPI: 89-114406/15.
PT      Hybridisation test for human papilloma virus in cell smears -
PT      by reaction with long labelled probe specific for particular
PT      virus types; esp. for examining cervical smears
PS      Disclosure; 39pp; English.
CC      The patent is for a rapid in situ hybridisation assay for detecting and
CC      typing human papilloma virus (HPV) in non-frozen cellular smears fixed to
CC      a support in absence of aldehyde-based crosslinking reagents. The assay

```

CC comprises: (1) combining nucleic acid in the sample with at least one  
 CC detectable probe able to hybridise with 1 or more HPV types; and (2)  
 CC detecting presence or absence of hybrid complexes. Opt. several probes  
 CC are used, eg one for HPV types 6 and 11, associated with benign warts,  
 CC and one for HPV types 16, 18, 31, 33 and 35, associated with cervical  
 CC cancer. The assay can differentiate between HPV types. It is esp. used  
 CC as a secondary test. The probes can be synthesised or cloned.  
 SQ Sequence 774 BP; 270 A; 132 C; 163 G; 209 T.

Query Match 100.0%; Score 10; DB 1; Length 774;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10  
 |||||  
 DB 603 TATGTTTAG 612

RESULT 8

005057 standard; DNA; 1975 BP.

DT 02-NOV-1990 (first entry)  
 DE Dral/Drai fragment of the patatin gene B33 promoter.  
 KW Expression cassette; transgenic potato plant; patatin gene B33;  
 KW tuber-specific regulatory region; plasmid pBI 101; beta-glucuronidase;  
 KW nopaline; terminator; ss.  
 OS Solanum tuberosum.  
 FH Key  
 FH Promoter 177..1713  
 FT Location/Qualifiers  
 FT /tag- a  
 FT /label-Drai/Drai promoter-fragment  
 FT /note-"from patatin gene B33"  
 FT 1738..1740  
 FT /tag- b  
 FT /function-translation start

EP-375092-A.  
 PN 27-JUN-1990.  
 PD 18-DEC-1989; 250117.  
 PR 21-DEC-1988; DE-843627.  
 PA (GENB-) INST GENBIOL FORSCH.  
 PI Roeha-Sosa M, Sonnewald U, Frommer WB, Willmitzer L, Stratman M;  
 DR WPI; 90-195269/26.  
 PT Tuber-specific expression cassette - for prodn. of transgenic  
 PT potato plants.  
 PS Disclosure; p; German.  
 PS 1.527 kb Dral/Drai fragment of the patatin B33 promoter (+14 to -1513)  
 was inserted into plasmid pBI 101, next to the coding region of  
 beta-glucuronidase and a poly-A containing nopaline-synthase terminator.  
 CC The resultant expression cassette pBI 101 B33 was transferred to an  
 CC agro bacterium (DBA 4404) which was used to infect potato leaves.  
 SQ Sequence 1975 BP; 690 A; 265 C; 310 G; 710 T;

Query Match 100.0%; Score 10; DB 1; Length 1975;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10  
 |||||  
 DB 334 TATGTTTAG 343

RESULT 9

N82025 standard; DNA; 1355 BP.

ID N82025.  
 AC N82025.  
 DT 22-OCT-1990 (first entry)  
 DE Fragment containing rat lipocortin gene  
 KW recombinant rat lipocortin; rat abdominal dropsy; ss.  
 OS Rattus.  
 FH Key  
 FH Location/Qualifiers

FT cds 28..1068  
 FT /tag- a  
 FT /product-lipocortin

PN J63276497-A.  
 PD 14-NOV-1988.  
 PF 08-MAY-1987; JP-112145.  
 PR 08-MAY-1987; J6-JP-112145.  
 PA (YAMA) Yamouchi Pharm KK.  
 DR WPI; 88-365616/51.  
 DR P-PSDB; P82062.

PT Recombinant rat lipocortin -  
 PT obd using gene derived from cells in rat abdominal dropsy  
 PS Disclosure; Japanese.  
 CC The coding sequence described in the features is specifically  
 CC claimed. mRNA was isolated from rat abdominal dropsy cells and cDNA  
 CC synthesised from it. Probes were synthesised according to the  
 CC partial amino acid sequence of rat lipocortin (see N82038 and  
 CC N82039). These were used to isolate plasmids contg the desired  
 CC sequence of lipocortin DNA.  
 CC See also P82063.  
 SQ Sequence 1355 BP; 433 A; 273 C; 302 G; 347 T;

Query Match 100.0%; Score 10; DB 1; Length 1355;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10  
 |||||  
 DB 1203 TATGTTTAG 1212

RESULT 10

N80859//C

ID N80859 standard; cDNA; 9643 BP.

AC N80859.

DT 15-NOV-1990 (first entry)  
 DE Sequence of entire HIV-2 ROD genome  
 KW LAV-II ROD; ss.  
 OS Human immunodeficiency virus ROD.

FH Key  
 FH Location/Qualifiers  
 FT cds 545..2110

FT /tag- a  
 FT /note-"p80801"  
 FT 1828..4908  
 FT /tag- b  
 FT /note-"p81773"

FT cds 4841..5485  
 FT /tag- c  
 FT /note-"p81774"

FT cds 5316..5654  
 FT /tag- d  
 FT /note-"p81775"

FT cds 5654..5968  
 FT /tag- e  
 FT /note-"p81776"

FT cds 5817..6113  
 FT /tag- f  
 FT /note-"p81777"

FT cds 6043..6153  
 FT /tag- g  
 FT /note-"p81778"

FT cds 6119..8692  
 FT /tag- h  
 FT /note-"p81779"

FT cds 8320..8372  
 FT /tag- i  
 FT /note-"p81780"

FT cds 8242..8508  
 FT /tag- j  
 FT /note-"p81781"

FT cds 8529..9299  
 FT /tag- k

FT cds

```

FT      /note="P81782"
PN      WO805440-A.
PD      28-JUL-1988.
PF      15-JAN-1988; F00025.
PR      15-APR-1987; FR-005398,
PA      (INSP) Inst Pasteur(Aliz/).
PI      Allison M, Montagnier L, Guetard D, Clavel F, Sonigo P, Guyader M,
PI      Tiollais P, Chakrabarti L, Desrosiers R;
DR      WPT-88-220290/31.
DR      P-PSDB: P80801, P81773, P81774, P81775, P81776, P81777, P81778, P81779,
DR      P81780, P81781, P81782.
PT      New peptide(s) with immunological properties of HIV-2 envelope protein -
PT      having the structure of simian immune deficiency virus proteins,
PS      useful in diagnosis and of vaccine components
PS      Disclosure: Fig 1A; 86pp; French.
CC      The SQ in n80859 was deposited on 21/2/86 at the CNCM under number I-522,
CC      reference name LAV-II ROD. It is the CDNA to HIV-2 ROD genomic RNA. SQ
CC      was compared with the SQ of the genome of SIV (Mac) (n80860) to identify
CC      common regions.
Sequence 9643 BP; 3296 A; 1969 C; 2399 G; 1979 T;

Query Match      100.0%; Score 10; DB 1; Length 9643;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TATGTTTAG 10
DB      644 TATGTTTAG 635

RESULT 11
ID      N80436/c
AC      N80436;
DT      16-DEC-1990 (first entry)
DE      Entire sequence of LAV EL I
KW      HIV; HIV-1; AIDS; diagnosis; vaccine; probe; hybridisation; ss.
OS      Lymphadenopathy associated virus EL I.
FH      Key
FT      Location/Qualifiers
FT      misc-feature
FT      1..98
FT      /tag= a
FT      /label=R
FT      99..182
FT      /tag= b
FT      /label=U5
FT      8683..9138
FT      /tag= c
FT      /label=U3
FT      9139..9236
FT      /tag= d
FT      /label=R
FT      336..1835
FT      /tag= e
FT      /label=GAG; P80884
FT      1634..4699
FT      /tag= f
FT      /label=POL; P81854
FT      4647..5222
FT      /tag= g
FT      /label=O; P81855
FT      5165..5452
FT      /tag= h
FT      /label=R; P81856
FT      5436..5651
FT      /tag= i
FT      /label=S; P81857
FT      5830..8388
FT      /tag= j
FT      /label=ENV; P81858
FT      8393..9010
FT      /tag= k
FT      /label=F; P81859

```

```

PN      WO8707906-A.
PD      30-DEC-1987.
PF      22-JUN-1987; E00326.
PR      23-JUN-1986; EP-401380.
PA      (INSP) Inst Pasteur.
PI      Alizon M, Sonigo P, Wain-Hobson S, Montagnier L;
DR      WPT-88-014396/02.
DR      P-PSDB: P80884, P81854, P81855, P81856, P81857, P81858, P81859.
PT      New variants of lymphadenopathy associated virus (LAV) -
PT      used for prodn. of DNA, antigens and antibodies used in
PT      diagnosis of AIDS and pre-AIDS
PS      Claim 3: Fig 7A-7J; 72pp; English.
CC      LAV EL I (n80436) and LAV MA L (n80437) were isolated from the peripheral
CC      blood lymphocytes of patients. The different AIDS virus isolates
CC      are designated by 3 letters of the patients name. Stable probes including
CC      the DNA sequences can be used for detection of the new LAV viruses or
CC      related viruses or DNA proviruses in eg biological samples. The proteins
CC      or peptides can be used for detection of antibodies induced in vivo and
CC      present in biological fluids. The DNA can also be used for the expression
CC      of LAV viral antigens for the prodn. of a vaccine against LAV. The
CC      polypeptides can also be used for the prodn. of antibodies for the
CC      detection of proteins related to the LAV viruses, partic. for diagnosis
CC      of AIDS or pre-AIDS.
Sequence 9236 BP; 3360 A; 1642 C; 2190 G; 2044 T;

Query Match      100.0%; Score 10; DB 1; Length 9236;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TATGTTTAG 10
DB      435 TATGTTTAG 426

RESULT 12
ID      N82198/c
AC      N82198;
DT      05-DEC-1990 (first entry)
DE      GAGROD sequence from HIV-2 ROD corresp. to the HIV-2 gag gene.
KW      Human immunodeficiency virus-2; HIV-2; AIDS; HIV-2 ROD; gag gene;
KW      env gene; GAGROD; p16; p26; p12; ss.
OS      HIV-2 ROD.
FH      EP-269520-A.
FT      01-JUN-1988.
FT      23-NOV-1987; 402631.
FT      21-NOV-1986; US-933184.
FT      22-JAN-1987; EP-400151.
FT      PA
FT      (INSP) Inst Pasteur.
FT      PI
FT      Montagnier L, Charetet S, Guetard D, Alizon M, Clavel F, Guyader M,
FT      Sonigo P, Brun-Vezinet F, Rey M, Rouzoulou C, Katlama C;
FT      WPT-88-149264/22.
FT      P-PSDB: P82676.
PT      New HIV-2 retrovirus causing AIDS - and new antigenic proteins,
PT      antibodies and complementary nucleic acid sequences.
PS      Claim 5; Page 29-30; 67pp; French.
CC      HIV-2 has RNA, esp. the gag and env regions, at least 50%, pref.
CC      70% homologous or even 90%, corresponding to the gag and env
CC      regions of HIV-2 ROD.
CC      GAGROD includes coding regions for p16:nucleotides 1-405;
CC      p26:nucleotides 406-1155;
CC      p12:nucleotides 1156-1566.
CC      HIV-2 retrovirus is infectious for human T4 lymphocytes and has the
CC      morphological and immunological properties of the strains CNCM I-502,
CC      -532, -642 and -643. HIV-2 causes AIDS and is distinct from HIV-1.
CC      The sequence can be used diagnostically (in hybridisation tests).
CC      See also N82199 for the ENVN sequence and N82200 for the
CC      R and U3 regions.
SQ      Sequence 1566 BP; 536 A; 338 C; 420 G; 272 T;

Query Match      100.0%; Score 10; DB 1; Length 1566;

```



Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10  
Db 100 TATGTTTAG 91

## RESULT 13

N70887/c  
ID N70887 standard; DNA; 1566 BP.  
AC N70887;  
DT 28-FEB-1991 (first entry)  
DE HIV virus gag sequence.  
KW Immunodeficiency virus; HIV-2; AIDS; ss.  
OS Homo sapiens.  
PN W08704459-A.  
PD 30-JUL-1987.  
PF 22-JAN-1987; 700025.  
PT 22-JAN-1986; FR-000911.  
PR 22-JAN-1986; FR-000910.  
PR 06-FEB-1986; FR-001635.  
PR 13-FEB-1986; FR-001985.  
PR 03-MAR-1986; US-835228.  
PR 18-MAR-1986; FR-003881.  
PR 24-MAR-1986; FR-004215.  
PR 06-OCT-1986; US-916080.  
PR 21-NOV-1986; US-933184.  
PR 22-JAN-1987; EP-400151.  
PA (INSP ) INST PASTEUR.  
PA (MONT/) MONTAGNIER L.  
PI Montagnier L, Charette S, Guetard D, Clavel F, Guyader M;  
PI Sonigop, Burnezeine F;  
DR WPI: 87-221261/31.  
DR P-PSDB; P70554.  
PT New type of human immuno-deficiency virus, infections for T4  
PT cells - and derived antigens, immunogens, monoclonal antibodies  
PT and nucleic acid sequences, eg. for diagnosis of AIDS  
PS Claim 5; Page 71; 116pp; French.  
CC The sequence encodes the gag protein of a new HIV-2 virus genome  
CC and is infectious for human T4-lymphocytes. Abs and probes may be  
CC raised, useful in HIV-2 vaccination and diagnosis respectively.  
CC See also N70886, N70886-89, P70550-53.  
CC Sequence 1566 BP; 536 A; 336 C; 420 G; 272 T;

Query Match 100.0%; Score 10; DB 1; Length 1566;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10  
Db 100 TATGTTTAG 91

## RESULT 14

Q10947  
ID Q10947 standard; DNA; 4355 BP.  
AC Q10947;  
DT 24-MAY-1991 (first entry)  
DE Bovine Coronavirus E2 gene.  
KW Bovine Coronavirus; BCoV; E2; E3; vaccine; ss.  
OS Bovine coronavirus - Quebec isolate.  
FH Key Location/Qualifiers  
FT cds 14..4102  
FT /\*tag= a  
FT /\*product= E2  
FT /\*tag= b  
FT /\*note= "conserved intergenic region"  
FT signal\_peptide 14..64  
FT /\*tag= c  
FT misc\_rna 3929..4027

FT /\*tag= d  
FT /label= transmembrane-domain

FN W09102752-A.  
PD 07-MAR-1991.  
PF 10-AUG-1990; CA0252.  
PR 22-AUG-1989; US-397689.  
PA (VETE-) VETERINARY INPECTIO.  
PI Parker MD, Cox GJ, Babluk LA;  
DR WPI: 91-087247/12.  
DR P-PSDB; R11061.  
PT Decyribonucleic acid encoding bovine coronavirus protein E2 and  
PT E3 - useful as vaccine component  
PS Disclosure, Fig 3 (1-6); 70pp; English.  
CC E2 is one of the four genes (N, E1, E2, E3) composing BCoV.  
CC Immediately 5' of the E2 gene on the viral genome is the E3 gene  
CC (Q10948). The E3 gene terminates 14 nucleotides upstream from  
CC the E2 polypeptide initiation codon.  
CC Coding sequences can either be prepared directly by synthetic  
CC methods based on this sequence, or by using the sequence to design  
CC oligonucleotide probes to clone coding sequences.  
CC The E2 and E3 polypeptides, or fragments of these, are used as  
CC subunit antigens in vaccines for protection against or ameliorating  
CC BCoV, without risk of infection.  
SO Sequence 4355 BP; 1219 A; 702 C; 852 G; 1582 T;

Query Match 100.0%; Score 10; DB 1; Length 4355;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10  
Db 164 TATGTTTAG 173

## RESULT 15

Q10093  
ID Q10093 standard; DNA; 35 BP.  
AC Q10093;  
DT 13-MAR-1991 (first entry)  
DE Probe 782 to Chlamydia trachomatis 16S rRNA.  
KW 16S rRNA; sexually transmitted diseases; ss.  
OS Synthetic.  
PN W09015159-A.  
PD 13-DEC-1990.  
PF 31-MAY-1990; U02989.  
PR 31-MAY-1989; US-359293.  
PA (GENE-) GENE-TRAK SYST.  
PI Shah J, Buharin A, Williams C, Mahan D, Lane DJ, King W;  
DR WPI: 91-007228/01.  
PT Nucleic acid probes - specific for r-RNA or r-DNA of Chlamydia  
PT trachomatis, useful in hybridisation assay for detection of e.g.  
PT sexual transmitted diseases  
PS Claim 3; Page 32; 35pp; English.  
CC The probe is one of nine 16S rRNA-targeted probes which are specific  
CC to Chlamydia trachomatis bacteria. It may be used in a variety of  
CC assay systems to detect e.g. respiratory, ocular or sexually  
CC transmitted diseases. See also Q10092 and Q10094-Q10110.  
SO Sequence 35 BP; 12 A; 7 C; 13 G; 13 T;

Query Match 100.0%; Score 10; DB 1; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10  
Db 7 TATGTTTAG 16

Search completed: September 18, 1999, 00:33:37  
Job time: 18977 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:03; Search time 2825.05 Seconds  
(without alignments)  
6.982 Million cell updates/sec

Title: US-08-956-518a-91  
Perfect score: 10  
Sequence: 1 TATGTTTAG 10

Scoring table: IDENTITY\_NDC

Searched: 2546578 seqs, 986266752 residues

abase :

EST.\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	10	100.0	208	D20858	D20858 HUMG501839
C 2	10	100.0	191	D30830	D30830 HUM11042 H
C 3	10	100.0	360	D35619	D35619 CELK02308F
C 4	10	100.0	259	D36779	D36779 CELK037A3F
C 5	10	100.0	339	M79084	M79084 EST01232 Su
C 6	10	100.0	403	M79539	M79539 WEST00096 M
C 7	10	100.0	509	M89400	M89400 CEL08B11 Ch
C 8	10	100.0	239	T02576	T02576 0213c3 CBSP
C 9	10	100.0	344	T03104	T03104 FB2484 Peta
C 10	10	100.0	336	T05062	T05062 EST02950 Fe
C 11	10	100.0	358	T05141	T05141 EST03029 Fe
C 12	10	100.0	385	T05247	T05247 EST03136 Fe
C 13	10	100.0	364	T09635	T09635 021lm3 gmbp
C 14	10	100.0	318	T10573	T10573 hpc156 Huma
C 15	10	100.0	324	T10574	T10574 hpc160 Huma
C 16	10	100.0	289	T11931	T11931 A929P Heart
C 17	10	100.0	374	T15644	T15644 IB1680 Infa
C 18	10	100.0	260	T16586	T16586 NIB1519 Nor
C 19	10	100.0	384	T17266	T17266 NIB608 Norm
C 20	10	100.0	285	T18987	T18987 b07030c Tes
C 21	10	100.0	349	T23480	T23480 seq3179 1-N
C 22	10	100.0	349	T23733	T23733 seq585 b4HB
C 23	10	100.0	292	T25848	T25848 EST209 BL29
C 24	10	100.0	552	T25873	T25873 ESTDIR93 CD
C 25	10	100.0	262	T28017	T28017 EST23952 Hu
C 26	10	100.0	302	T29131	T29131 EST70246 Hu
C 27	10	100.0	268	T29634	T29634 EST88290 Hu
C 28	10	100.0	316	T29708	T29708 EST90996 Hu
C 29	10	100.0	272	T30096	T30096 EST112333 R
C 30	10	100.0	288	T30100	T30100 EST112343 H
C 31	10	100.0	204	T30294	T30294 EST14376 Hu
C 32	10	100.0	325	T32724	T32724 EST53631 Hu
C 33	10	100.0	345	T36036	T36036 EST95994 Hu
C 34	10	100.0	389	T37196	T37196 EST102276 S
C 35	10	100.0	370	T37394	T37394 EST102507 S
C 36	10	100.0	313	T38155	T38155 EST103532 S
C 37	10	100.0	339	T40472	T40472 Y802B01 .s2
C 38	10	100.0	384	T47530	T47530 Y15805 .r1
C 39	10	100.0	354	T50974	T50974 YD71B01 .s1
C 40	10	100.0	360	T51037	T51037 YD71B01 .r1
C 41	10	100.0	390	T54364	T54364 Y91504 .s3
C 42	10	100.0	382	T54798	T54798 YD42B01 .s1
C 43	10	100.0	258	T55438	T55438 YD38B06 .s1
C 44	10	100.0	182	T20143	T20143 H5AABMIG T
C 45	10	100.0	806	HSW009483	AL044633 Homo sapi

#### ALIGNMENTS

RESULT 1  
LOCUS D20858 208 bp mRNA  
DEFINITION HUMG501839 Human promyelocyte Homo sapiens cDNA clone mp0405 3',  
ACCESSION D20858  
NID 9304678  
VERSION D20858.1 GI:504678

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Homiidae; Homo.  
TITLE 1 (bases 1 to 208)  
Muraoka, K., Matsubara, K., Fukushima, A., Yoshii, J. and Okubo, K.  
Chromosomal assignments of 3'-directed partial cDNA sequences  
representing novel genes expressed in granulocytoid cells  
Genomics 23, 379-389 (1994)

JOURNAL MEDLINE 95137584  
REFERENCE 2 (bases 1 to 208)  
AUTHORS Okubo, K., Fukushima, A., Yoshii, J., Miyama, T., Kojima, Y.,  
Yoshinari, H., Arimoto, J. and Matsubara, K.  
Gene expression of human promyelocytic cell line HL60 before and  
after induction of differentiation. A new application of 3'-directed  
cDNA sequencing  
Unpublished (1993)

JOURNAL COMMENT  
Contact: Okubo, K., Fukushima, A., Yoshii, J., Miyama, T., Kojima, Y.,  
Yoshinari, H., Arimoto, J. and Matsubara, K.  
Institute for Molecular and Cellular Biology  
Osaka University  
3-1 Yamada-Oka, Suita, Osaka 565, Japan.  
Location/Qualifiers  
1..208  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="7"  
/clone="mp0405"  
/clone\_1lb="Human promyelocyte"  
/note="Female, adult, cell\_line = HL60, cell\_type =  
promyelocyte." 28 c 20 g 71 t 2 others

BASE COUNT 87 a 28 c 20 g 71 t 2 others  
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 208;  
Best Local Similarity 100.0%; Pred. No. 2.1e+04;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10  
|||||  
DB 128 TATGTTTAG 119

LOCUS 130/c  
DEFINITION D30830 191 bp mRNA EST 08-FEB-1995  
HUMH1042 Human fetal lung Homo sapiens cDNA 5', mRNA sequence.  
ACCESSION D30830  
NID 9643710  
VERSION D30830.1 GI:643710  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Homiidae; Homo.  
TITLE 1 (bases 1 to 191)  
Sudo, K., Chinen, K. and Nakamura, Y.  
2058 expressed sequence tags (ESTs) from a human fetal lung cDNA  
library  
Genomics 24, 276-279 (1995)  
JOURNAL MEDLINE 95213017  
COMMENT Contact: Yusuke Nakamura  
Institute of Medical Science  
University of Tokyo  
4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan  
Tel: 81-3-5449-5372  
Fax: 81-3-5449-5433  
Email: yusuke@ims.u-tokyo.ac.jp.  
Location/Qualifiers

FEATURES

source 1..191  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1lb="Human fetal lung"

BASE COUNT 54 a 49 c 24 g 54 t 10 others  
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 191;  
Best Local Similarity 100.0%; Pred. No. 2.2e+04;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10  
|||||  
DB 90 TATGTTTAG 81

RESULT 3  
D35619/c 360 bp mRNA EST 08-AUG-1994  
LOCUS CELK02308F Yui Kohara unpublished cDNA Caenorhabditis elegans cDNA  
DEFINITION clone yk2308 5', mRNA sequence.  
ACCESSION D35619  
NID 9527069  
VERSION D35619.1 GI:527069  
KEYWORDS EST.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans  
Eukaryota: Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabditida;  
Rhabditina; Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.  
1 (bases 1 to 360)  
Kohara, Y., Mituoka, H., Nishigaki, A., Motomachi, T., Sugimoto, A. and  
Tabara, H.  
Toward an expression map of the C. elegans genome  
Unpublished (1994)

TITLE  
JOURNAL  
COMMENT Contact: Yui Kohara  
Gene Library Lab  
National Institute of Genetics  
Yata 111, Mishima, Shizuoka 411, Japan  
Tel: 0559-75-0771  
Fax: 0559-75-6240  
Email: ykohara@dbj.nig.ac.jp.  
Location/Qualifiers  
1..360  
/organism="Caenorhabditis elegans"  
/strain="CBI489 him-8(e1489)"  
/db\_xref="taxon:6239"  
/clone="yk2308"  
/clone\_1lb="Yui Kohara unpublished cDNA"  
/note="dev\_stage=varied, sex=Hermaphrodite male,  
tissue\_type=whole animal" 68 c 71 g 84 t 3 others

BASE COUNT 134 a 68 c 71 g 84 t 3 others  
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 360;  
Best Local Similarity 100.0%; Pred. No. 1.7e+04;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10  
|||||  
DB 336 TATGTTTAG 327

RESULT 4  
D36779/c 259 bp mRNA EST 08-AUG-1994  
LOCUS CELK037A3F Yui Kohara unpublished cDNA Caenorhabditis elegans cDNA  
DEFINITION clone yk37a3 5', mRNA sequence.  
ACCESSION D36779  
NID 9528695  
VERSION D36779.1 GI:528695

KEYWORDS EST.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.  
REFERENCE Rhabditina; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae.  
AUTHORS 1 (bases 1 to 259)  
Kohara, Y., Mitsukawa, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and Tabara, H.  
TITLE Toward an expression map of the C. elegans genome  
JOURNAL Unpublished (1994)  
COMMENT Contact: Yuji Kohara  
Gene Library Lab  
National Institute of Genetics  
Yata 111, Mishima, Shizuoka 411, Japan  
Tel: 0559-75-0771  
Fax: 0559-75-6240  
Email: ykoha@nig.ac.jp.  
FEATURES Location/Qualifiers  
1. 259  
/organism="Caenorhabditis elegans"  
/strain="CB1489 him-8(e1489)"  
/db\_xref="taxon:6239"  
/clone="YK37a3"  
/note="dev-stage=varied, sex=Hermaphrodite male, tissue\_type=whole animal"

BASE COUNT 111 a 48 c 26 g 73 t 1 others  
ORIGIN

Query Match  
Best Local Similarity 100.0%; Score 10; DB 20; Length 259;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10  
|||||  
Db 225 TATGTTTAG 216

RESULT 5  
LOCUS M79084 339 bp mRNA EST 26-MAY-1992  
DEFINITION EST01232 Subtracted Hippocampus, Strata gene (cat. #936205) Homo  
ACCESSION M79084 sapiens cDNA clone HHCPR27, mRNA sequence.  
M79084  
M79084 9273397  
M79084.1 GI:273397  
EST.  
WORDS human.  
JRCCE  
ORGANISM Homo sapiens  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlavet@igrr.org  
Seq primer: M13 Forward.  
LOCATION/Qualifiers  
1. 339  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):78612"  
/db\_xref="GDB:D052053E"

FEATURES  
SOURCE  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlavet@igrr.org  
Seq primer: M13 Forward.  
LOCATION/Qualifiers  
1. 339  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):78612"  
/db\_xref="GDB:D052053E"

/db\_xref="taxon:9606"  
/clone="HHCPR27"  
/clone="HHCPR27"  
#936205"  
/note="Vector: LambdaZAP-II; The hippocampus library (#4 above) was subtracted with a fibroblast cell line cDNA library (Stratagene cat. #936209; W138 lung fibroblast cell line; Oligo-dT + random primed cDNA synthesis; LambdaZAP-II vector, 1.0kb average insert size.) by the method of Sive & St. John (Nucleic Acids Res. 16:10937, 1988)."

BASE COUNT 115 a 66 c 66 g 92 t  
ORIGIN

Query Match  
Best Local Similarity 100.0%; Score 10; DB 20; Length 339;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10  
|||||  
Db 181 TATGTTTAG 190

RESULT 6  
LOCUS M79559 403 bp mRNA EST 30-JUN-1992  
DEFINITION WEST00096 Mixed stage, Strata gene (cat. #937006) Caenorhabditis  
ACCESSION M79559 elegans cDNA clone CEMS03 similar to Cytochrome b, mitochondrial,  
M79559 mRNA sequence.  
M79559 9271582  
M79559.1 GI:271582  
EST.  
WORDS human.  
JRCCE  
ORGANISM Caenorhabditis elegans.  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlavet@igrr.org  
Seq primer: M13 Forward.  
LOCATION/Qualifiers  
1. 403  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/clone="CEMS03"  
/clone="lib="Mixed stage, Strata gene (cat. #937006)"  
/note="Vector: Uni-ZAP XR; C. elegans mixed stage cDNA library. Stratagene catalog #937006. The library is oligo dT primed and directionally cloned in the Uni-ZAP XR vector."

REFERENCE AUTHORS  
McCormick, W.R., Adams, M.D., Kelley, J.M., Fitzgerald, M.G., Fields, C.  
Caenorhabditis elegans expressed sequence tags identify gene families and potential disease gene homologues  
Nature Genet. 1, 124-131 (1992)  
93250983

JOURNAL MEDLINE  
COMMENT

CONTACT: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlavet@igrr.org  
Seq primer: M13 Forward.  
LOCATION/Qualifiers  
1. 403  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/clone="CEMS03"  
/clone="lib="Mixed stage, Strata gene (cat. #937006)"  
/note="Vector: Uni-ZAP XR; C. elegans mixed stage cDNA library. Stratagene catalog #937006. The library is oligo dT primed and directionally cloned in the Uni-ZAP XR vector."

BASE COUNT 111 a 43 c 81 g 164 t 4 others  
ORIGIN

Query Match  
Best Local Similarity 100.0%; Score 10; DB 20; Length 403;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10  
 |||||||  
 Db 15 TATGTTTAG 24

RESULT 7  
 M89400 509 bp mRNA EST 02-DEC-1992  
 LOCUS M89400  
 DEFINITION CE108B11 Chris Martin sorted cDNA library Caenorhabditis elegans  
 CDNA clone cm08b11 5' similar to cytochrome b homologous peptide,  
 mRNA sequence.  
 ACCSSION M89400  
 NID 9275295  
 VERSION M89400.1 GI:275295  
 KEYWORDS EST  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;  
 Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 509)  
 WATERSTON R., MARTIN C., CRAXTON M., HUYNH C., COULSON A.,  
 HILLIER L., DURBIN R.K., GREEN P., SHOWNKEEN R., HALLIFRAN N.,  
 HAWKINS T., WILSON R., BERRS M., DU Z., THOMAS K., THIERRY-MEG J.,  
 and SULSTON J.  
 A survey of expressed genes in Caenorhabditis elegans  
 NATURE GENET. 1, 114-123 (1992)  
 93250982

TITLE JOURNAL  
 MEDLINE  
 COMMENT

REFERENCE  
 AUTHORS  
 WATERSTON R., MARTIN C., CRAXTON M., HUYNH C., COULSON A.,  
 HILLIER L., DURBIN R.K., GREEN P., SHOWNKEEN R., HALLIFRAN N.,  
 HAWKINS T., WILSON R., BERRS M., DU Z., THOMAS K., THIERRY-MEG J.,  
 and SULSTON J.  
 A survey of expressed genes in Caenorhabditis elegans  
 NATURE GENET. 1, 114-123 (1992)  
 93250982

COMMENT  
 Contact: Waterston R.H.(USA) and Sulston J.E.(UK)  
 (USA) Dept. of Genetics or (UK)  
 (USA) Washington Univ. School of Medicine or (UK) MRC Laboratory of  
 Molecular Biology  
 Box 8233, 4566 Scott Ave., St. Louis, MI 63110, USA, or Hills  
 Road, Cambridge CB2 2QH, UK  
 Tel: (USA) (314)3627072 or (UK) (0223)248011  
 Fax: (USA) (314)3624137 or (UK) (0223)402008  
 Email: tw@ematoled.wustl.edu or jes@mc-lmba.cambridge.ac.uk  
 single read.

FEATURES  
 Location/Qualifiers  
 1..509  
 /organism="Caenorhabditis elegans"  
 /strain="Bristol N2"  
 /db\_xref="taxon:6239"  
 /clone="cm08b11"  
 /clone\_lib="Chris Martin sorted cDNA library"  
 /lab\_host="MC1061"  
 /note="Vector: lambda phage SHLX2; Mixed stage  
 hermaprodite cDNA library. Partially normalized by  
 successively picking groups of clones that didn't  
 hybridize to previously picked clones. Vector: lambda phage  
 SHLX2 (Lipshitz, D.H. et al., Gene 88:25-36 (1990)). Host:  
 MC1061."  
 MCI061." 49 c 96 g 204 t 9 others

BASE COUNT 151 a 49 c 96 g 204 t 9 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 509;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10  
 |||||||  
 Db 365 TATGTTTAG 374

RESULT 8  
 T02576 239 bp mRNA EST 26-MAY-1992  
 LOCUS T02576  
 DEFINITION C01303 cbSPFHB3.1, Debopam Chakrabarti Plasmodium falciparum CDNA  
 Clone PF0213C, mRNA sequence.  
 ACCSSION T02576  
 NID 9271430  
 VERSION T02576.1 GI:271430

KEYWORDS EST.  
 SOURCE malaria parasite P. falciparum.  
 ORGANISM Plasmodium falciparum  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 1 (bases 1 to 239)  
 CHAKRABARTI D., REDDY G.R., DAME J.B., ALMITRA E.C., LAPLIS P.J.,  
 FERL R.J., YANG T.P., ROWE T.C., and SCHUSTER S.M.  
 Analysis of expressed sequence tags from Plasmodium falciparum  
 Mol. Biochem. Parasitol. 66, 97-104 (1994)  
 95075403

TITLE JOURNAL  
 MEDLINE  
 COMMENT

REFERENCE  
 AUTHORS  
 CHAKRABARTI D., REDDY G.R., DAME J.B., ALMITRA E.C., LAPLIS P.J.,  
 FERL R.J., YANG T.P., ROWE T.C., and SCHUSTER S.M.  
 Analysis of expressed sequence tags from Plasmodium falciparum  
 Mol. Biochem. Parasitol. 66, 97-104 (1994)  
 95075403

COMMENT  
 Contact: Dame JB  
 Department of Pathobiology  
 University of Florida  
 2015 SW 16th Ave., Bldg. 1017, Gainesville, FL 32611-0880  
 Tel: 352 392 4700  
 Fax: 352 392 9704  
 Email: dame@icbr.ufl.edu  
 Seq primer: 73.

FEATURES  
 Location/Qualifiers  
 1..239  
 /organism="Plasmodium falciparum"  
 /db\_xref="taxon:5833"  
 /clone="PF0213C"  
 /clone\_lib="cbSPFHB3.1, Debopam Chakrabarti"  
 /lab\_host="E. coli XL1-Blue"  
 /note="Vector: Bluescript SK(+); PolyA+ RNA, from  
 asynchronous blood stage parasites of the cloned Honduran  
 HB3 isolate cultured in vitro, was reverse transcribed  
 using an Oligo dt-Xho I primer. Second strand was prepared  
 using RNase H and DNA polymerase I. EcoR I adapters were  
 ligated to the cDNA, and it was cleaved with Xho I.  
 Prepared fragments were ligated into EcoR I + Xho I  
 double-digested Bluescript SK(+), and transformed E. coli  
 XL1-Blue."  
 98 a 20 c 25 g 84 t 12 others

BASE COUNT 98 a 20 c 25 g 84 t 12 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 239;  
 Best Local Similarity 100.0%; Pred. No. 2e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10  
 |||||||  
 Db 217 TATGTTTAG 208

RESULT 9  
 T03104 344 bp mRNA EST 14-FEB-1997  
 LOCUS T03104  
 DEFINITION FB24B4 Fetal brain, Strata gene Homo sapiens CDNA clone FB24B4  
 3' end, mRNA sequence.  
 T03104  
 ACCSSION T03104  
 NID 9314345  
 VERSION T03104.1 GI:314345  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 344)  
 KHAN A.S., WILCOX A.S., POLYMERPOPOULOS M.H., HOPKINS J.A.,  
 STEVENS T.J., ROBINSON M., ORPANA A.K., and SIKELA J.M.  
 Single pass sequencing and physical and genetic mapping of human  
 CDNA  
 NATURE GENET. 2, 180-185 (1992)  
 94258200  
 On Sep 21, 1992 this sequence version replaced gi:279243.

TITLE JOURNAL  
 MEDLINE  
 COMMENT

REFERENCE  
 AUTHORS  
 KHAN A.S., WILCOX A.S., POLYMERPOPOULOS M.H., HOPKINS J.A.,  
 STEVENS T.J., ROBINSON M., ORPANA A.K., and SIKELA J.M.  
 Single pass sequencing and physical and genetic mapping of human  
 CDNA  
 NATURE GENET. 2, 180-185 (1992)  
 94258200

COMMENT  
 Contact: Sikele JM  
 Department of Pharmacology  
 University of Colorado Health Sciences Center

Box C236, 4200 E. 9th Ave, Denver CO 80262-0236  
Tel: 3032708637  
Fax: 3032707097  
Email: nikri@etally.uchsc.edu

Seq primer: -21M13 Universal.  
Location/Qualifiers

## FEATURES

source

1.344

/organism="Homo sapiens"

/db\_xref="ATCC (Inhost):86749"

/db\_xref="GDB:D052872E"

/db\_xref="taxon:9606"

/clone="FB24B4"

/clone.lib="Fetal brain, Stratagene"

/note="Vector: Lambda ZAP II; The FB library (catalog #937201, Stratagene) was constructed by directional cloning and oligo d(T)-priming in Lambda ZAP II phage utilizing mRNA from a 17 week human fetus (total brain)."

98 a 75 c 76 g 95 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 10; DB 20; Length 344;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10

Db 184 TATGTTTAG 175

RESULT 10

LOCUS

T05062 336 bp mRNA EST 30-JUN-1993

DEFINITION EST02950 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA

clone HFB055 similar to EST containing 0 family repeat, mRNA

sequence.

T05062 336 bp mRNA EST 30-JUN-1993

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

T05062.1 GI:316215

Best Local Similarity 100.0%; Pred. No. 1.7e+04;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10

Db 167 TATGTTTAG 176

RESULT 11

LOCUS

T05141 358 bp mRNA EST 30-JUN-1993

DEFINITION EST03029 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA

clone HFB055 similar to EST containing 0 family repeat, mRNA

sequence.

T05141 358 bp mRNA EST 30-JUN-1993

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

T05141.1 GI:316293

AUTHORS Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.  
 TITLE 3,400 expressed sequence tags identify diversity of transcripts  
 JOURNAL Nature Genet. 4, 256-267 (1993)  
 MEDLINE 93364420  
 COMMENT

Contact: Adams, MD  
 The Institute for Genomic Research  
 932 Clopper Road, Gaithersburg, MD 20878  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: mdadams@tigr.org

## FEATURES

source Location/Qualifiers

1..385  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (lnhost):81973"  
 /db\_xref="taxon:9606"  
 /clone="HFBCT86"  
 /note="Vector: Fetal brain, Striatum (cat#936206)"  
 oligo-dt + random primed cDNA synthesis; lambdaZAP-II  
 vector, 1.0kb average insert size."  
 BASE COUNT 94 a 76 c 81 g 133 t 1 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 385;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10  
 |||||  
 Db 133 TATGTTTAG 124

RESULT 13  
 T09635 364 bp DNA EST 29-NOV-1993  
 LOCUS 021lm3 gmbpFHB3.1, G. Roman Reddy Plasmodium falciparum genomic  
 DEFINITION clone 021lm, mRNA sequence.  
 ACCESSION T09635  
 T09635  
 NID 9319467  
 VERSION T09635.1 GI:319467  
 KEYWORDS EST.  
 SOURCE malaria parasite P. falciparum.  
 ORGANISM Plasmodium falciparum  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 1 (bases 1 to 364)  
 Reddy,G.R., Chakrabarti,D., Schuster,S.M., Ferl,R.J., Almitra,E.C.  
 and Dame,J.B.  
 Gene sequence tags from Plasmodium falciparum genomic DNA fragments  
 prepared by the genease activity of mung bean nuclease  
 Proc. Natl. Acad. Sci. U.S.A. 90, 9867-9871 (1993)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT  
 Contact: Dame JB  
 Department of Pathobiology  
 University of Florida  
 2015 SW 16th Ave., Bldg. 1017, Gainesville, FL 32611-0880  
 Tel: 352 392 4700  
 Fax: 352 392 9704  
 Email: dame@icbr.tifas.ufl.edu  
 Seq primer: 73.

## FEATURES

source

Location/Qualifiers  
 1..364  
 /organism="Plasmodium falciparum"  
 /db\_xref="taxon:5833"  
 /clone="021lm"  
 /clone\_1lb="gmbpFHB3.1, G. Roman Reddy"  
 /lab\_host="E. coli XL1-Blue"  
 /note="Vector: pBluescript SK(+); genomic DNA, from  
 asynchronous blood stage parasites of the cloned Honduran

HB3 isolate cultured in vitro, was digested with mung bean  
 nuclease in the presence of 308 formamide at 500C  
 (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988.  
 Nucleic Acids Research 16:6883-6896). The ends of the  
 fragments were ligated using T4 DNA polymerase, and the  
 fragments were ligated to EcoR V-cleaved and  
 dephosphorylated pBluescript SK(+). Recombinant plasmids  
 transformed E. coli XL1-Blue.  
 BASE COUNT 94 a 61 c 51 g 151 t 7 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 364;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10  
 |||||  
 Db 48 TATGTTTAG 57

RESULT 14  
 T10573/c 318 bp mRNA EST 29-NOV-1993  
 LOCUS hbc156 Human pancreatic islet Homo sapiens cDNA clone hbc156 5'end,  
 DEFINITION mRNA sequence.  
 ACCESSION T10573  
 T10573  
 NID 9390727  
 VERSION T10573.1 GI:390727  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 318)  
 Takeda,J., Yano,H., Eng,S., Zeng,Y. and Bell,G.I.  
 A molecular inventory of human pancreatic islets: sequence analysis  
 of 1000 cDNA clones  
 Hum. Mol. Genet. 2, 1793-1798 (1993)  
 JOURNAL 94108427  
 MEDLINE  
 COMMENT

Contact: Bell GI or Takeda J  
 HHMI  
 Univ. of Chicago  
 5841 S. Maryland Ave., MC1028, Chicago IL 60637  
 Tel: 3127029116  
 Fax: 3127020271  
 Email: g-bell@uchicago.edu  
 Seq primer: SK primer.  
 Location/Qualifiers

FEATURES  
 source 1..318

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="hbc156"  
 /clone\_1lb="Human pancreatic islet"  
 /note="Vector: lambda ZAPIT, Site\_1: Eco RI; Site\_2: Xho  
 I; mRNA was prepared from normal adult human islets. cDNA  
 was directionally synthesized from the Xho I in the vector  
 to the EcoRI site. cDNA was size fractionated to remove  
 sequences <1000 bp in size."  
 BASE COUNT 114 a 66 c 58 g 80 t  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 318;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10  
 |||||  
 Db 154 TATGTTTAG 145

RESULT 15  
 T10574/c 324 bp mRNA EST 29-NOV-1993  
 LOCUS hbc160 Human pancreatic islet Homo sapiens cDNA clone hbc160 5' end,  
 DEFINITION mRNA sequence.  
 ACCESSION T10574  
 NID 9380728  
 VERSION T10574.1 GI:390728  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 324)  
 AUTHORS Takeda, J., Yano, H., Eng, S., Zeng, Y. and Bell, G.I.  
 TITLE A molecular inventory of human pancreatic islets: sequence analysis  
 of 1000 cDNA clones  
 JOURNAL Hum. Mol. Genet. 2, 1793-1798 (1993)  
 MEDLINE 94108427  
 COMMENT

Contact: Bell GI or Takeda J  
 HHMI  
 Univ. of Chicago  
 5841 S. Maryland Ave., MC1028, Chicago IL 60637  
 Tel: 312/7029116  
 Fax: 312/7020271  
 Email: g-bell@uchicago.edu  
 Seq primer: SK primer.  
 Location/Qualifiers

FEATURES  
 source

1..324  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="hbc160"  
 /clone\_id="Human pancreatic islet"  
 /note="Vector: Lambda ZAPIT; Site\_1: Eco RI; Site\_2: Xho  
 I; mRNA was prepared from normal adult human islets. cDNA  
 was directionally synthesized from the Xho I in the vector  
 to the EcoRI site. cDNA was size fractionated to remove  
 sequences <1000 bp in size."  
 BASE COUNT 115 a 67 c 63 g 79 t  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 324;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TATGTTTAG 10  
 |||||  
 158 TATGTTTAG 149

Search completed: September 17, 1999, 21:28:05  
 Job time: 14269 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:33:37 ; Search time 425.19 seconds

(without alignments)  
5.884 Million cell updates/sec

Title: US-08-956-518a-92

Sequence: 1 CTTCTCCACAG 10

Scoring table: IDENTITY\_NUC

Database: 311585 seqs, 125096042 residues

Database: N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	100.0	7377	1 Q03005	Full length T4 CDN
2	10	100.0	6151	1 Q03006	Full length T4 CDN
3	10	100.0	660	1 N90493	A DNA sequence of
4	10	100.0	1310	1 N90334	C-terminus of the
5	10	100.0	614	1 N90502	Heavy chain antilbo
6	10	100.0	919	1 N90695	DNA sequence of po
7	10	100.0	595	1 N90696	DNA sequence of po
8	10	100.0	527	1 N91649	Reshaped light cha
9	10	100.0	150	1 N91481	Start of coding re
10	10	100.0	540	1 N91645	Heavy chain of mon
11	10	100.0	443	1 N91820	DNA sequence of th
12	10	100.0	616	1 N91463	Sequence of human
13	10	100.0	600	1 N93069	Fragment of PCR-el
14	10	100.0	5518	1 N90647	Nucleotide sequenc
15	10	100.0	5566	1 N90648	Nucleotide sequenc
16	10	100.0	5413	1 N90649	Nucleotide sequenc
17	10	100.0	5365	1 N90646	Nucleotide sequenc
18	10	100.0	608	1 N90500	Light chain antilbo
19	10	100.0	765	1 Q04040	Anti-Leu 3a heavy
20	10	100.0	4565	1 Q04077	Human 32k alveolar
21	10	100.0	2720	1 Q04081	Vector pASPCg-SV1
22	10	100.0	2761	1 Q04082	Vector pMT-Apo:GHS
23	10	100.0	4024	1 Q04027	Human angiotensin
24	10	100.0	497	1 Q04283	Modified human lym
25	10	100.0	7316	1 Q04555	Plasmid p170.2 enc
26	10	100.0	527	1 Q03290	New synthetic huma
27	10	100.0	134525	1 Q04525	Total base sequenc
28	10	100.0	620	1 Q04462	M13 VPCRI with 19
29	10	100.0	818	1 Q04461	M13 VPCRI with 19
30	10	100.0	4219	1 Q04786	Sequence encoding
31	10	100.0	4353	1 Q04787	Sequence encoding
32	10	100.0	972	1 Q05105	Sequence encoding
33	10	100.0	592	1 N80631	Bovine Bone Morpho
34	10	100.0	2500	1 N80632	Human Bone Morpho
35	10	100.0	3850	1 N81564	Sequence of new pl
36	10	100.0	2839	1 Q05513	1177 CDNA encoding
37	10	100.0	664	1 Q05516	Upstream sequence
38	10	100.0	646	1 Q05517	Upstream sequence
39	10	100.0	641	1 Q05518	Upstream sequence
40	10	100.0	664	1 Q05519	Upstream sequence
41	10	100.0	7377	1 Q05607	Plasmid p170.2 inc
42	10	100.0	6151	1 Q05608	Plasmid pBG381 inc
43	10	100.0	6928	1 Q06161	Part 1 of CDNA der

c 44 10 100.0 345 1 Q07017  
45 10 100.0 6614 1 X26905

HIV-1 Rev M6 mutan  
6.6 kb transcript

## ALIGNMENTS

RESULT 1  
ID 003005  
AC 003005 standard; CDNA; 7377 BP.  
DE 29-MAY-1990 (first entry)  
DT Full length T4 CDNA of plasmid p170-2.  
KW Soluble T4; p170-2; anti-retroviral agent; AIDS; ARC; HIV; AZT.  
PN M08911860-A.  
PD 14-DEC-1989.  
PF 08-JUN-1989; U02453.  
PR 10-JUN-1988; US-204645.  
PS (GEO-) General Hospital Corp.  
PT (BIOT) Biogen Inc.  
DR Fisher RA, Schooley RT, Hirsch MS, Johnson VA, Walker BD;  
P-PSDB; R04031.  
PT Combinations of soluble T4 protein and anti-retroviral agent -  
PT having synergistic activity in treatment and prevention of AIDS,  
PS Disclosure; h.  
CC Soluble T4 constructs may be produced by truncating this sequence, to  
CC remove the transmembrane and intracytoplasmic domains whilst retaining  
CC the extracellular region responsible for HIV binding. The sol. T4 is  
CC combined with an anti-viral agent such as AZT.  
SQ See also Q03006.  
Sequence 7377 BP; 1760 A; 1954 C; 1902 G; 1761 T;

Query Match 100.0%; Score 10; DB 1; Length 7377;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTCTCCACAG 10  
DB 962 CTTCTCCACAG 971

RESULT 2  
ID 003006  
AC 003006 standard; CDNA; 6151 BP.  
DE 29-MAY-1990 (first entry)  
DT Full length T4 CDNA of plasmid pBG381.  
KW Soluble T4; pBG381; anti-retroviral agent; AIDS; ARC; HIV; AZT.  
PN M08911860-A.  
PD 14-DEC-1989.  
PF 08-JUN-1989; U02453.  
PR 10-JUN-1988; US-204645.  
PS (GEO-) General Hospital Corp.  
PT (BIOT) Biogen Inc.  
DR Fisher RA, Schooley RT, Hirsch MS, Johnson VA, Walker BD;  
P-PSDB; R04032.  
PT Combinations of soluble T4 protein and anti-retroviral agent -  
PT having synergistic activity in treatment and prevention of AIDS,  
PS Disclosure; h.  
CC Soluble T4 constructs may be produced by truncating this sequence, to  
CC remove the transmembrane and intracytoplasmic domains whilst retaining  
CC the extracellular region responsible for HIV binding. The sol. T4 is  
CC combined with an anti-viral agent such as AZT.  
SQ See also Q03005.  
Sequence 6151 BP; 1493 A; 1589 C; 1586 G; 1483 T;

Query Match 100.0%; Score 10; DB 1; Length 6151;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10  
 |||||  
 DB 970 CTCCTCCACAG 979

RESULT 3  
 N90493  
 ID N90493 standard; DNA; 660 BP.  
 AC N90493;  
 DT 20-OCT-1989 (first entry)  
 DE A DNA sequence of a VH region gene.  
 KW VH region gene; Immunoglobulin; HIV.  
 OS Mus musculus  
 FH Key Location/Qualifiers  
 FT exon 156..201  
 FT exon 278..646  
 FT exon /tag= a  
 FT exon /tag= b

PN EP-327000-A.  
 PD 09-AUG-1989.  
 PE 30-JAN-1989; 101583.  
 PR 30-JAN-1988; JP-20255.  
 PR 08-JUL-1988; JP-171385.  
 PA (KAGA) The Chemo-Sero-Therapeutic Research Institute.  
 PI Maeda H, Eda Y, Kimachi K, Tokiyoshi S, Matsushita S, Hattori T,  
 PI Takatsuki K;  
 DR P-PSDB; P90542.  
 PT Chimeric anti-human immune virus antibodies - contg. mouse variable  
 PT regions and human constant regions for diagnosis, treatment and  
 PT prevention of AIDS.  
 PS Disclosure; Fig 5; 33pp; English.  
 CC The DNA is a VH region gene. See P90542, N90491-2, and N90495.  
 SO Sequence 660 BP; 184 A; 155 C; 144 G; 177 T;

Query Match 100.0%; Score 10; DB 1; Length 660;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10  
 |||||  
 DB 268 CTCCTCCACAG 277

JULT 4  
 N90334/C  
 ID N90334 standard; DNA; 1310 BP.  
 AC N90334;  
 DT 1-NOV-1989 (first entry)  
 DE C-terminus of the superoxide dismutase C-100 fusion polypeptide  
 KW Hepatitis C virus; CDNA; fusion polypeptide C100-3; superoxide dismutase;  
 KM Non-A, non-B hepatitis.  
 OS Pan troglodytes  
 FH Key Location/Qualifiers  
 FT cds 1..1166  
 FT /tag= a  
 FT /tag= b  
 FT /tag= c  
 FT /tag= d  
 FT /tag= e  
 FT /tag= f  
 FT /tag= g  
 FT /tag= h  
 FT /tag= i  
 FT /tag= j  
 FT /tag= k  
 FT /tag= l  
 FT /tag= m  
 FT /tag= n  
 FT /tag= o  
 FT /tag= p  
 FT /tag= q  
 FT /tag= r  
 FT /tag= s  
 FT /tag= t  
 FT /tag= u  
 FT /tag= v  
 FT /tag= w  
 FT /tag= x  
 FT /tag= y  
 FT /tag= z  
 FT /tag= aa  
 FT /tag= ab  
 FT /tag= ac  
 FT /tag= ad  
 FT /tag= ae  
 FT /tag= af  
 FT /tag= ag  
 FT /tag= ah  
 FT /tag= ai  
 FT /tag= aj  
 FT /tag= ak  
 FT /tag= al  
 FT /tag= am  
 FT /tag= an  
 FT /tag= ao  
 FT /tag= ap  
 FT /tag= aq  
 FT /tag= ar  
 FT /tag= as  
 FT /tag= at  
 FT /tag= au  
 FT /tag= av  
 FT /tag= aw  
 FT /tag= ax  
 FT /tag= ay  
 FT /tag= az  
 FT /tag= ba  
 FT /tag= bb  
 FT /tag= bc  
 FT /tag= bd  
 FT /tag= be  
 FT /tag= bf  
 FT /tag= bg  
 FT /tag= bh  
 FT /tag= bi  
 FT /tag= bj  
 FT /tag= bk  
 FT /tag= bl  
 FT /tag= bm  
 FT /tag= bn  
 FT /tag= bo  
 FT /tag= bp  
 FT /tag= bq  
 FT /tag= br  
 FT /tag= bs  
 FT /tag= bt  
 FT /tag= bu  
 FT /tag= bv  
 FT /tag= bw  
 FT /tag= bx  
 FT /tag= by  
 FT /tag= bz  
 FT /tag= ca  
 FT /tag= cb  
 FT /tag= cc  
 FT /tag= cd  
 FT /tag= ce  
 FT /tag= cf  
 FT /tag= cg  
 FT /tag= ch  
 FT /tag= ci  
 FT /tag= cj  
 FT /tag= ck  
 FT /tag= cl  
 FT /tag= cm  
 FT /tag= cn  
 FT /tag= co  
 FT /tag= cp  
 FT /tag= cq  
 FT /tag= cr  
 FT /tag= cs  
 FT /tag= ct  
 FT /tag= cu  
 FT /tag= cv  
 FT /tag= cw  
 FT /tag= cx  
 FT /tag= cy  
 FT /tag= cz  
 FT /tag= da  
 FT /tag= db  
 FT /tag= dc  
 FT /tag= dd  
 FT /tag= de  
 FT /tag= df  
 FT /tag= dg  
 FT /tag= dh  
 FT /tag= di  
 FT /tag= dj  
 FT /tag= dk  
 FT /tag= dl  
 FT /tag= dm  
 FT /tag= dn  
 FT /tag= do  
 FT /tag= dp  
 FT /tag= dq  
 FT /tag= dr  
 FT /tag= ds  
 FT /tag= dt  
 FT /tag= du  
 FT /tag= dv  
 FT /tag= dw  
 FT /tag= dx  
 FT /tag= dy  
 FT /tag= dz  
 FT /tag= ea  
 FT /tag= eb  
 FT /tag= ec  
 FT /tag= ed  
 FT /tag= ee  
 FT /tag= ef  
 FT /tag= eg  
 FT /tag= eh  
 FT /tag= ei  
 FT /tag= ej  
 FT /tag= ek  
 FT /tag= el  
 FT /tag= em  
 FT /tag= en  
 FT /tag= eo  
 FT /tag= ep  
 FT /tag= eq  
 FT /tag= er  
 FT /tag= es  
 FT /tag= et  
 FT /tag= eu  
 FT /tag= ev  
 FT /tag= ew  
 FT /tag= ex  
 FT /tag= ey  
 FT /tag= ez  
 FT /tag= fa  
 FT /tag= fb  
 FT /tag= fc  
 FT /tag= fd  
 FT /tag= fe  
 FT /tag= ff  
 FT /tag= fg  
 FT /tag= fh  
 FT /tag= fi  
 FT /tag= fj  
 FT /tag= fk  
 FT /tag= fl  
 FT /tag= fm  
 FT /tag= fn  
 FT /tag= fo  
 FT /tag= fp  
 FT /tag= fq  
 FT /tag= fr  
 FT /tag= fs  
 FT /tag= ft  
 FT /tag= fu  
 FT /tag= fv  
 FT /tag= fw  
 FT /tag= fx  
 FT /tag= fy  
 FT /tag= fz  
 FT /tag= ga  
 FT /tag= gb  
 FT /tag= gc  
 FT /tag= gd  
 FT /tag= ge  
 FT /tag= gf  
 FT /tag= gg  
 FT /tag= gh  
 FT /tag= gi  
 FT /tag= gj  
 FT /tag= gk  
 FT /tag= gl  
 FT /tag= gm  
 FT /tag= gn  
 FT /tag= go  
 FT /tag= gp  
 FT /tag= gq  
 FT /tag= gr  
 FT /tag= gs  
 FT /tag= gt  
 FT /tag= gu  
 FT /tag= gv  
 FT /tag= gw  
 FT /tag= gx  
 FT /tag= gy  
 FT /tag= gz  
 FT /tag= ha  
 FT /tag= hb  
 FT /tag= hc  
 FT /tag= hd  
 FT /tag= he  
 FT /tag= hf  
 FT /tag= hg  
 FT /tag= hi  
 FT /tag= hj  
 FT /tag= hk  
 FT /tag= hl  
 FT /tag= hm  
 FT /tag= hn  
 FT /tag= ho  
 FT /tag= hp  
 FT /tag= hq  
 FT /tag= hr  
 FT /tag= hs  
 FT /tag= ht  
 FT /tag= hu  
 FT /tag= hv  
 FT /tag= hw  
 FT /tag= hx  
 FT /tag= hy  
 FT /tag= hz  
 FT /tag= ia  
 FT /tag= ib  
 FT /tag= ic  
 FT /tag= id  
 FT /tag= ie  
 FT /tag= if  
 FT /tag= ig  
 FT /tag= ih  
 FT /tag= ii  
 FT /tag= ij  
 FT /tag= ik  
 FT /tag= il  
 FT /tag= im  
 FT /tag= in  
 FT /tag= io  
 FT /tag= ip  
 FT /tag= iq  
 FT /tag= ir  
 FT /tag= is  
 FT /tag= it  
 FT /tag= iu  
 FT /tag= iv  
 FT /tag= iw  
 FT /tag= ix  
 FT /tag= iy  
 FT /tag= iz  
 FT /tag= ja  
 FT /tag= jb  
 FT /tag= jc  
 FT /tag= jd  
 FT /tag= je  
 FT /tag= jf  
 FT /tag= jg  
 FT /tag= jh  
 FT /tag= ji  
 FT /tag= jj  
 FT /tag= jk  
 FT /tag= jl  
 FT /tag= jm  
 FT /tag= jn  
 FT /tag= jo  
 FT /tag= jp  
 FT /tag= jq  
 FT /tag= jr  
 FT /tag= js  
 FT /tag= jt  
 FT /tag= ju  
 FT /tag= jv  
 FT /tag= jw  
 FT /tag= jx  
 FT /tag= jy  
 FT /tag= jz  
 FT /tag= ka  
 FT /tag= kb  
 FT /tag= kc  
 FT /tag= kd  
 FT /tag= ke  
 FT /tag= kf  
 FT /tag= kg  
 FT /tag= kh  
 FT /tag= ki  
 FT /tag= kj  
 FT /tag= kl  
 FT /tag= km  
 FT /tag= kn  
 FT /tag= ko  
 FT /tag= kp  
 FT /tag= kq  
 FT /tag= kr  
 FT /tag= ks  
 FT /tag= kt  
 FT /tag= ku  
 FT /tag= kv  
 FT /tag= kw  
 FT /tag= kx  
 FT /tag= ky  
 FT /tag= kz  
 FT /tag= la  
 FT /tag= lb  
 FT /tag= lc  
 FT /tag= ld  
 FT /tag= le  
 FT /tag= lf  
 FT /tag= lg  
 FT /tag= lh  
 FT /tag= li  
 FT /tag= lj  
 FT /tag= lk  
 FT /tag= ll  
 FT /tag= lm  
 FT /tag= ln  
 FT /tag= lo  
 FT /tag= lp  
 FT /tag= lq  
 FT /tag= lr  
 FT /tag= ls  
 FT /tag= lt  
 FT /tag= lu  
 FT /tag= lv  
 FT /tag= lw  
 FT /tag= lx  
 FT /tag= ly  
 FT /tag= lz  
 FT /tag= ma  
 FT /tag= mb  
 FT /tag= mc  
 FT /tag= md  
 FT /tag= me  
 FT /tag= mf  
 FT /tag= mg  
 FT /tag= mh  
 FT /tag= mi  
 FT /tag= mj  
 FT /tag= mk  
 FT /tag= ml  
 FT /tag= mm  
 FT /tag= mn  
 FT /tag= mo  
 FT /tag= mp  
 FT /tag= mq  
 FT /tag= mr  
 FT /tag= ms  
 FT /tag= mt  
 FT /tag= mu  
 FT /tag= mv  
 FT /tag= mw  
 FT /tag= mx  
 FT /tag= my  
 FT /tag= mz  
 FT /tag= na  
 FT /tag= nb  
 FT /tag= nc  
 FT /tag= nd  
 FT /tag= ne  
 FT /tag= nf  
 FT /tag= ng  
 FT /tag= nh  
 FT /tag= ni  
 FT /tag= nj  
 FT /tag= nk  
 FT /tag= nl  
 FT /tag= nm  
 FT /tag= nn  
 FT /tag= no  
 FT /tag= np  
 FT /tag= nq  
 FT /tag= nr  
 FT /tag= ns  
 FT /tag= nt  
 FT /tag= nu  
 FT /tag= nv  
 FT /tag= nw  
 FT /tag= nx  
 FT /tag= ny  
 FT /tag= nz  
 FT /tag= oa  
 FT /tag= ob  
 FT /tag= oc  
 FT /tag= od  
 FT /tag= oe  
 FT /tag= of  
 FT /tag= og  
 FT /tag= oh  
 FT /tag= oi  
 FT /tag= oj  
 FT /tag= ok  
 FT /tag= ol  
 FT /tag= om  
 FT /tag= on  
 FT /tag= oo  
 FT /tag= op  
 FT /tag= oq  
 FT /tag= or  
 FT /tag= os  
 FT /tag= ot  
 FT /tag= ou  
 FT /tag= ov  
 FT /tag= ow  
 FT /tag= ox  
 FT /tag= oy  
 FT /tag= oz  
 FT /tag= pa  
 FT /tag= pb  
 FT /tag= pc  
 FT /tag= pd  
 FT /tag= pe  
 FT /tag= pf  
 FT /tag= pg  
 FT /tag= ph  
 FT /tag= pi  
 FT /tag= pj  
 FT /tag= pk  
 FT /tag= pl  
 FT /tag= pm  
 FT /tag= pn  
 FT /tag= po  
 FT /tag= pp  
 FT /tag= pq  
 FT /tag= pr  
 FT /tag= ps  
 FT /tag= pt  
 FT /tag= pu  
 FT /tag= pv  
 FT /tag= pw  
 FT /tag= px  
 FT /tag= py  
 FT /tag= pz  
 FT /tag= qa  
 FT /tag= qb  
 FT /tag= qc  
 FT /tag= qd  
 FT /tag= qe  
 FT /tag= qf  
 FT /tag= qg  
 FT /tag= qh  
 FT /tag= qi  
 FT /tag= qj  
 FT /tag= qk  
 FT /tag= ql  
 FT /tag= qm  
 FT /tag= qn  
 FT /tag= qo  
 FT /tag= qp  
 FT /tag= qq  
 FT /tag= qr  
 FT /tag= qs  
 FT /tag= qt  
 FT /tag= qu  
 FT /tag= qv  
 FT /tag= qw  
 FT /tag= qx  
 FT /tag= qy  
 FT /tag= qz  
 FT /tag= ra  
 FT /tag= rb  
 FT /tag= rc  
 FT /tag= rd  
 FT /tag= re  
 FT /tag= rf  
 FT /tag= rg  
 FT /tag= rh  
 FT /tag= ri  
 FT /tag= rj  
 FT /tag= rk  
 FT /tag= rl  
 FT /tag= rm  
 FT /tag= rn  
 FT /tag= ro  
 FT /tag= rp  
 FT /tag= rq  
 FT /tag= rr  
 FT /tag= rs  
 FT /tag= rt  
 FT /tag= ru  
 FT /tag= rv  
 FT /tag= rw  
 FT /tag= rx  
 FT /tag= ry  
 FT /tag= rz  
 FT /tag= sa  
 FT /tag= sb  
 FT /tag= sc  
 FT /tag= sd  
 FT /tag= se  
 FT /tag= sf  
 FT /tag= sg  
 FT /tag= sh  
 FT /tag= si  
 FT /tag= sj  
 FT /tag= sk  
 FT /tag= sl  
 FT /tag= sm  
 FT /tag= sn  
 FT /tag= so  
 FT /tag= sp  
 FT /tag= sq  
 FT /tag= sr  
 FT /tag= ss  
 FT /tag= st  
 FT /tag= su  
 FT /tag= sv  
 FT /tag= sw  
 FT /tag= sx  
 FT /tag= sy  
 FT /tag= sz  
 FT /tag= ta  
 FT /tag= tb  
 FT /tag= tc  
 FT /tag= td  
 FT /tag= te  
 FT /tag= tf  
 FT /tag= tg  
 FT /tag= th  
 FT /tag= ti  
 FT /tag= tj  
 FT /tag= tk  
 FT /tag= tl  
 FT /tag= tm  
 FT /tag= tn  
 FT /tag= to  
 FT /tag= tp  
 FT /tag= tq  
 FT /tag= tr  
 FT /tag= ts  
 FT /tag= tt  
 FT /tag= tu  
 FT /tag= tv  
 FT /tag= tw  
 FT /tag= tx  
 FT /tag= ty  
 FT /tag= tz  
 FT /tag= ua  
 FT /tag= ub  
 FT /tag= uc  
 FT /tag= ud  
 FT /tag= ue  
 FT /tag= uf  
 FT /tag= ug  
 FT /tag= uh  
 FT /tag= ui  
 FT /tag= uj  
 FT /tag= uk  
 FT /tag= ul  
 FT /tag= um  
 FT /tag= un  
 FT /tag= uo  
 FT /tag= up  
 FT /tag= uq  
 FT /tag= ur  
 FT /tag= us  
 FT /tag= ut  
 FT /tag= uu  
 FT /tag= uv  
 FT /tag= uw  
 FT /tag= ux  
 FT /tag= uy  
 FT /tag= uz  
 FT /tag= va  
 FT /tag= vb  
 FT /tag= vc  
 FT /tag= vd  
 FT /tag= ve  
 FT /tag= vf  
 FT /tag= vg  
 FT /tag= vh  
 FT /tag= vi  
 FT /tag= vj  
 FT /tag= vk  
 FT /tag= vl  
 FT /tag= vm  
 FT /tag= vn  
 FT /tag= vo  
 FT /tag= vp  
 FT /tag= vq  
 FT /tag= vr  
 FT /tag= vs  
 FT /tag= vt  
 FT /tag= vu  
 FT /tag= vv  
 FT /tag= vw  
 FT /tag= vx  
 FT /tag= vy  
 FT /tag= vz  
 FT /tag= wa  
 FT /tag= wb  
 FT /tag= wc  
 FT /tag= wd  
 FT /tag= we  
 FT /tag= wf  
 FT /tag= wg  
 FT /tag= wh  
 FT /tag= wi  
 FT /tag= wj  
 FT /tag= wk  
 FT /tag= wl  
 FT /tag= wm  
 FT /tag= wn  
 FT /tag= wo  
 FT /tag= wp  
 FT /tag= wq  
 FT /tag= wr  
 FT /tag= ws  
 FT /tag= wt  
 FT /tag= wu  
 FT /tag= wv  
 FT /tag= ww  
 FT /tag= wx  
 FT /tag= wy  
 FT /tag= wz  
 FT /tag= xa  
 FT /tag= xb  
 FT /tag= xc  
 FT /tag= xd  
 FT /tag= xe  
 FT /tag= xf  
 FT /tag= xg  
 FT /tag= xh  
 FT /tag= xi  
 FT /tag= xj  
 FT /tag= xk  
 FT /tag= xl  
 FT /tag= xm  
 FT /tag= xn  
 FT /tag= xo  
 FT /tag= xp  
 FT /tag= xq  
 FT /tag= xr  
 FT /tag= xs  
 FT /tag= xt  
 FT /tag= xu  
 FT /tag= xv  
 FT /tag= xw  
 FT /tag= xx  
 FT /tag= xy  
 FT /tag= xz  
 FT /tag= ya  
 FT /tag= yb  
 FT /tag= yc  
 FT /tag= yd  
 FT /tag= ye  
 FT /tag= yf  
 FT /tag= yg  
 FT /tag= yh  
 FT /tag= yi  
 FT /tag= yj  
 FT /tag= yk  
 FT /tag= yl  
 FT /tag= ym  
 FT /tag= yn  
 FT /tag= yo  
 FT /tag= yp  
 FT /tag= yq  
 FT /tag= yr  
 FT /tag= ys  
 FT /tag= yt  
 FT /tag= yu  
 FT /tag= yv  
 FT /tag= yw  
 FT /tag= yx  
 FT /tag= yy  
 FT /tag= yz  
 FT /tag= za  
 FT /tag= zb  
 FT /tag= zc  
 FT /tag= zd  
 FT /tag= ze  
 FT /tag= zf  
 FT /tag= zg  
 FT /tag= zh  
 FT /tag= zi  
 FT /tag= zj  
 FT /tag= zk  
 FT /tag= zl  
 FT /tag= zm  
 FT /tag= zn  
 FT /tag= zo  
 FT /tag= zp  
 FT /tag= zq  
 FT /tag= zr  
 FT /tag= zs  
 FT /tag= zt  
 FT /tag= zu  
 FT /tag= zv  
 FT /tag= zw  
 FT /tag= zx  
 FT /tag= zy  
 FT /tag= zz

PA (CHIR-) Chiron Corporation.  
 PI Houghton M, Choo QL, Kuo G;  
 DR WPI; 89-215054/30.  
 DR P-PSDB; P90182.  
 PT Hepatitis C virus gene - used for prodn. of polynucleotide probes,  
 PT polypeptide(s) and antibodies for diagnosis, prevention and treatment  
 PT of infection  
 PS Disclosure; Fig 36; 235pp; English.  
 CC The sequence shows the sequence of the C-terminus of the superoxide  
 CC dismutase (SOD)-C100 fusion polypeptide. The sequence encodes antigens  
 CC which react with antibodies in patients with non-A non-B hepatitis  
 CC (NANBH). The CDNA can be used to design probes, or to synthesise  
 CC polypeptides, which are used to diagnose HCV-induced NANBH, to  
 CC raise antibodies for immunoassay or treatment, or to produce vaccines.  
 CC Misc. feature b shows the SOD C-terminal; c shows an adaptor; d shows the  
 CC NANBH polypeptide; and e shows an extra sequence. See also P90182,  
 CC N90303-33, and N90335-6.  
 SO Sequence 1310 BP; 262 A; 388 C; 378 G; 282 T;

Query Match 100.0%; Score 10; DB 1; Length 1310;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10  
 |||||  
 DB 1176 CTCCTCCACAG 1167

RESULT 5  
 N90502  
 ID N90502 standard; DNA; 614 BP.  
 AC N90502;  
 DT 16-NOV-1989 (first entry)  
 DE Heavy chain antibody variable domain; complementary determining region;  
 KW Heavy chain variable domain; lymphoma; immunosuppression.  
 OS Rattus norvegicus  
 FH Key Location/Qualifiers  
 FT misc\_feature 48  
 FT /tag= a  
 FT /tag= b  
 FT /tag= c  
 FT /tag= d  
 FT /tag= e  
 FT /tag= f  
 FT /tag= g  
 FT /tag= h  
 FT /tag= i  
 FT /tag= j  
 FT /tag= k  
 FT /tag= l  
 FT /tag= m  
 FT /tag= n  
 FT /tag= o  
 FT /tag= p  
 FT /tag= q  
 FT /tag= r  
 FT /tag= s  
 FT /tag= t  
 FT /tag= u  
 FT /tag= v  
 FT /tag= w  
 FT /tag= x  
 FT /tag= y  
 FT /tag= z  
 FT /tag= aa  
 FT /tag= ab  
 FT /tag= ac  
 FT /tag= ad  
 FT /tag= ae  
 FT /tag= af  
 FT /tag= ag  
 FT /tag= ah  
 FT /tag= ai  
 FT /tag= aj  
 FT /tag= ak  
 FT /tag= al  
 FT /tag= am  
 FT /tag= an  
 FT /tag= ao  
 FT /tag= ap  
 FT /tag= aq  
 FT /tag= ar  
 FT /tag= as  
 FT /tag= at  
 FT /tag= au  
 FT /tag= av  
 FT /tag= aw  
 FT /tag= ax  
 FT /tag= ay  
 FT /tag= az  
 FT /tag= ba  
 FT /tag= bb  
 FT /tag= bc  
 FT /tag= bd  
 FT /tag= be  
 FT /tag= bf  
 FT /tag= bg  
 FT /tag= bh  
 FT /tag= bi  
 FT /tag= bj  
 FT /tag= bk  
 FT /tag= bl  
 FT /tag= bm  
 FT /tag= bn  
 FT /tag= bo  
 FT /tag= bp  
 FT /tag= bq  
 FT /tag= br  
 FT /tag= bs  
 FT /tag= bt  
 FT /tag= bu  
 FT /tag= bv  
 FT /tag= bw  
 FT /tag= bx  
 FT /tag= by  
 FT /tag= bz  
 FT /tag= ca  
 FT /tag= cb  
 FT /tag= cc  
 FT /tag= cd  
 FT /tag= ce  
 FT /tag= cf  
 FT /tag= cg  
 FT /tag= ch  
 FT /tag= ci  
 FT /tag= cj  
 FT /tag= ck  
 FT /tag= cl  
 FT /tag= cm  
 FT /tag= cn  
 FT /tag= co  
 FT /tag= cp  
 FT /tag= cq  
 FT /tag= cr  
 FT /tag= cs  
 FT /tag= ct  
 FT /tag= cu  
 FT /tag= cv  
 FT /tag= cw  
 FT /tag= cx  
 FT /tag= cy  
 FT /tag= cz  
 FT /tag= da  
 FT /tag= db  
 FT /tag= dc  
 FT /tag= dd  
 FT /tag= de  
 FT /tag= df  
 FT /tag= dg  
 FT /tag= dh  
 FT /tag= di  
 FT /tag= dj  
 FT /tag= dk  
 FT /tag= dl  
 FT /tag= dm  
 FT /tag= dn  
 FT /tag= do  
 FT /tag= dp  
 FT /tag= dq  
 FT /tag= dr  
 FT /tag= ds  
 FT /tag= dt  
 FT /tag= du  
 FT /tag= dv  
 FT /tag= dw  
 FT /tag= dx  
 FT /tag= dy  
 FT /tag= dz  
 FT /tag= ea  
 FT /tag= eb  
 FT /tag= ec  
 FT /tag= ed  
 FT /tag= ee  
 FT /tag= ef  
 FT /tag= eg  
 FT /tag= eh  
 FT /tag= ei  
 FT /tag= ej  
 FT /tag= ek  
 FT /tag= el  
 FT /tag= em  
 FT /tag= en  
 FT /tag= eo  
 FT /tag= ep  
 FT /tag= eq  
 FT /tag= er  
 FT /tag= es  
 FT /tag= et  
 FT /tag= eu  
 FT /tag= ev  
 FT /tag= ew  
 FT /tag= ex  
 FT /tag= ey  
 FT /tag= ez  
 FT /tag= fa  
 FT /tag= fb  
 FT /tag= fc  
 FT /tag= fd  
 FT /tag= fe  
 FT /tag= ff  
 FT /tag= fg  
 FT /tag= fh  
 FT /tag= fi  
 FT /tag= fj  
 FT /tag= fk  
 FT /tag= fl  
 FT /tag= fm  
 FT /tag= fn  
 FT /tag= fo  
 FT /tag= fp  
 FT /tag= fq  
 FT /tag= fr  
 FT /tag= fs  
 FT /tag= ft  
 FT /tag= fu  
 FT /tag= fv  
 FT /tag= fw  
 FT /tag= fx  
 FT /tag= fy  
 FT /tag= fz  
 FT /tag= ga  
 FT /tag= gb  
 FT /tag= gc  
 FT /tag= gd  
 FT /tag= ge  
 FT /tag= gf  
 FT /tag= gg  
 FT /tag= gh  
 FT /tag= gi  
 FT /tag= gj  
 FT /tag= gk  
 FT /tag= gl  
 FT /tag= gm  
 FT /tag= gn  
 FT /tag= go  
 FT /tag= gp  
 FT /tag= gq  
 FT /tag= gr  
 FT /tag= gs  
 FT /tag= gt  
 FT /tag= gu  
 FT /tag= gv  
 FT /tag= gw  
 FT /tag= gx  
 FT /tag= gy  
 FT /tag= gz  
 FT /tag= ha  
 FT /tag= hb  
 FT /tag= hc  
 FT /tag= hd  
 FT /tag= he  
 FT /tag= hf  
 FT /tag= hg  
 FT /tag= hi  
 FT /tag= hj  
 FT /tag= hk  
 FT /tag= hl  
 FT /tag= hm  
 FT /tag= hn  
 FT /tag= ho  
 FT /tag= hp  
 FT /tag= hq  
 FT /tag= hr  
 FT /tag= hs  
 FT /tag= ht  
 FT /tag= hu  
 FT /tag= hv  
 FT /tag= hw  
 FT /tag= hx  
 FT /tag= hy  
 FT /tag= hz  
 FT /tag= ia  
 FT /tag= ib  
 FT /tag= ic  
 FT /tag= id  
 FT /tag= ie  
 FT /tag= if  
 FT /tag= ig  
 FT /tag= ih  
 FT /tag= ii  
 FT /tag= ij  
 FT /tag= ik  
 FT /tag= il  
 FT /tag= im  
 FT /tag= in  
 FT /tag= io  
 FT /tag= ip  
 FT /tag= iq  
 FT /tag= ir  
 FT /tag= is  
 FT /tag= it  
 FT /tag= iu  
 FT /tag= iv  
 FT /tag= iw  
 FT /tag= ix  
 FT /tag= iy  
 FT /tag= iz  
 FT /tag= ja  
 FT /tag= jb  
 FT /tag= jc  
 FT /tag= jd  
 FT /tag= je  
 FT /tag= jf  
 FT /tag= jg  
 FT /tag= jh  
 FT /tag= ji  
 FT /tag= jj  
 FT /tag= jk  
 FT /tag= jl  
 FT /tag= jm  
 FT /tag= jn  
 FT /tag= jo  
 FT /tag= jp  
 FT /tag= jq  
 FT /tag= jr  
 FT /tag= js  
 FT /tag= jt  
 FT /tag= ju  
 FT /tag= jv  
 FT /tag= jw  
 FT /tag= jx  
 FT /tag= jy  
 FT /tag= jz  
 FT /tag= ka  
 FT /tag= kb  
 FT /tag= kc  
 FT /tag= kd  
 FT /tag= ke  
 FT /tag= kf  
 FT /tag= kg  
 FT /tag= kh  
 FT /tag= ki  
 FT /tag= kj  
 FT /tag= kl  
 FT /tag= km  
 FT /tag= kn  
 FT /tag= ko  
 FT /tag= kp  
 FT /tag= kq  
 FT /tag= kr  
 FT /tag= ks  
 FT /tag= kt  
 FT /tag= ku  
 FT /tag= kv  
 FT /tag= kw  
 FT /tag= kx  
 FT /tag= ky  
 FT /tag= kz  
 FT /tag= la  
 FT /tag= lb  
 FT /tag= lc  
 FT /tag= ld  
 FT /tag= le  
 FT /tag= lf  
 FT /tag= lg  
 FT /tag= lh  
 FT /tag= li  
 FT /tag= lj  
 FT /tag= lk  
 FT /tag= ll  
 FT /tag= lm  
 FT /tag= ln  
 FT /tag= lo  
 FT /tag= lp  
 FT /tag= lq  
 FT /tag= lr  
 FT /tag= ls  
 FT /tag= lt  
 FT /tag= lu  
 FT /tag= lv  
 FT /tag= lw  
 FT /tag= lx  
 FT /tag= ly  
 FT /tag= lz  
 FT /tag= ma  
 FT /tag= mb  
 FT /tag= mc  
 FT /tag= md  
 FT /tag= me  
 FT /tag= mf  
 FT /tag= mg  
 FT /tag= mh  
 FT /tag= mi  
 FT /tag= mj  
 FT /tag= mk  
 FT /tag= ml  
 FT /tag= mm  
 FT /tag= mn  
 FT /tag= mo  
 FT /tag= mp  
 FT /tag= mq  
 FT /tag= mr  
 FT /tag= ms  
 FT /tag= mt  
 FT /tag= mu  
 FT /tag= mv  
 FT /tag= mw  
 FT /tag= mx  
 FT /tag= my  
 FT /tag= mz  
 FT /tag= na  
 FT /tag= nb  
 FT /tag= nc  
 FT /tag= nd  
 FT /tag= ne  
 FT /tag= nf  
 FT /tag= ng  
 FT /tag= nh  
 FT /tag= ni  
 FT /tag= nj  
 FT /tag= nk  
 FT /tag= nl  
 FT /tag= nm  
 FT /tag= nn  
 FT /tag= no  
 FT /tag= np  
 FT /tag= nq  
 FT /tag= nr  
 FT /tag= ns  
 FT /tag= nt  
 FT /tag= nu  
 FT /tag= nv  
 FT /tag= nw  
 FT /tag= nx  
 FT /tag= ny  
 FT /tag= nz  
 FT /tag= oa  
 FT /tag= ob  
 FT /tag= oc  
 FT /tag= od  
 FT /tag= oe  
 FT /tag= of  
 FT /tag= og  
 FT /tag= oh  
 FT /tag= oi  
 FT /tag= oj  
 FT /tag= ok  
 FT /tag= ol  
 FT /tag= om  
 FT /tag= on  
 FT /tag= oo  
 FT /tag= op  
 FT /tag= oq  
 FT /tag= or  
 FT /tag= os  
 FT /tag= ot  
 FT /tag= ou  
 FT /tag= ov  
 FT /tag= ow  
 FT /tag= ox  
 FT /tag= oy  
 FT /tag= oz  
 FT /tag= pa  
 FT /tag= pb  
 FT /tag= pc  
 FT /tag= pd  
 FT /tag= pe  
 FT /tag= pf  
 FT /tag= pg  
 FT /tag= ph  
 FT /tag= pi  
 FT /tag= pj  
 FT /tag= pk  
 FT /tag= pl  
 FT /tag= pm  
 FT /tag= pn  
 FT /tag= po  
 FT /tag= pp  
 FT /tag= pq  
 FT /tag= pr  
 FT /tag= ps  
 FT /tag= pt  
 FT /tag= pu  
 FT /tag= pv  
 FT /tag= pw  
 FT /tag= px  
 FT /tag= py  
 FT /tag= pz  
 FT /tag= qa  
 FT /tag= qb  
 FT /tag= qc  
 FT /tag= qd  
 FT /tag= qe  
 FT /tag= qf  
 FT /tag= qg  
 FT /tag= qh  
 FT /tag= qi  
 FT /tag= qj  
 FT /tag= qk  
 FT /tag= ql  
 FT /tag= qm  
 FT /tag= qn  
 FT /tag= qo  
 FT /tag= qp  
 FT /tag= qq  
 FT /tag= qr  
 FT /tag= qs  
 FT /tag= qt  
 FT /tag= qu  
 FT /tag= qv  
 FT /tag= qw  
 FT /tag= qx  
 FT /tag= qy  
 FT /tag= qz  
 FT /tag= ra  
 FT /tag= rb  
 FT /tag= rc  
 FT /tag= rd

PT Antibody which binds to antigen Campath-1 - having complementary  
PT determining region of rat origin and foreign variable domain framework  
PS Disclousure; fig 2a; 30pp; English.  
CC The sequence encodes rat immunoglobulin heavy chain (see P90917).  
CC Misc features a and b show where RNA begins, h and l encode signal  
CC peptides, and j, k and l encode complementary determining regions 1, 2  
CC and 3 resp.  
SQ Sequence 614 BP; 164 A; 156 C; 153 G; 141 T;

Query Match  
Best Local Similarity 100.0%; Score 10; DB 1; Length 614;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCACAG 10  
|||||||  
Db 228 CTCCTCACAG 237

RESULT 6  
N90695  
ID N90695 standard; DNA: 919 BP.  
AC N90695;  
DT 15-JUL-1990 (first entry)  
DE DNA sequence of portion of plasmid pF8CIS containing cytomegalovirus  
DE enhancer, promoter and splice donor site and Ig variable region  
KW plasmid pF8CIS; cytomegalovirus; enhancer; promoter;  
KW stabilising sequence; splice donor intron sequence; Ig variable region;  
KW splice acceptor sequence; human embryonic kidney cells (293); JM2.  
FH Cytomegalovirus +  
FT Key  
FT Location/Qualifiers  
FT Promoter  
FT 1..732  
FT /tag= a  
FT /note="Cytomegalovirus enhancer, promoter."  
FT 733..900  
FT /tag= b  
FT /note="stabilising sequence, ie. splice donor intron  
FT sequence, the Ig variable region intron and splice  
FT acceptor sequence."  
FT misc\_rna  
FT 621  
FT /tag= c  
FT /note="Begin RNA"  
FT 800..900  
FT /tag= d  
FT /note="IG synthetic 100mer"  
FT 897..916  
FT /tag= e  
FT /note="PstI-ClaI converter"  
FT misc\_feature  
FT 897..916

FT EP-309235-A.  
PN 29-MAR-1989.  
PD 22-SEP-1988; 308784.  
PF 25-SEP-1987; US-101712.  
PR (GUTH) Genentech Inc.  
PI Gorman CM;  
DR WPI; 89-095394/13.  
PT Transient expression system for recombinant proteins -  
PT encoding trans-activating host cell transfected with vector  
PS Disclousure; 44 pp; English.  
CC Expression vector pF8CIS contains the cytomegalovirus enhancer and prom-  
CC oter, the cytomegalovirus splice donor site and a portion of an intron,  
CC the Ig variable region intron and splice acceptor site, the cDNA encoding  
CC factor VIII and the SV40 polyadenylation site. It is used as a method for  
CC prodn. of a desired heterologous protein in a eukaryotic host cell.  
CC The method comprises: transfecting a eukaryotic host cell with a vector  
CC encoding trans-activating protein; transfecting the host cell with  
CC an expression vector having a sequence of a ds DNA comprising, a  
CC stabilising sequence downstream of a promoter and upstream of a DNA  
CC encoding the amino acid sequence of the desired protein, and a poly-  
CC -adenylation sequence upstream to a transcription termination site.  
CC The pref. host cells are human embryonic kidney cells (293) or JM2.  
CC The method provides useful quantities of a desired protein in a

CC relatively short period of time without having to establish continuous  
CC prodn. The transient expression system optimises the interaction between  
CC specific vector components and certain trans-activating proteins.  
SQ Sequence 919 BP; 219 A; 244 C; 223 G; 233 T;

Query Match  
Best Local Similarity 100.0%; Score 10; DB 1; Length 919;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCACAG 10  
|||||||  
Db 867 CTCCTCACAG 876

RESULT 7  
N90696  
ID N90696 standard; DNA: 595 BP.  
AC N90696;  
DT 15-JUL-1990 (first entry)  
DE DNA sequence of portion of plasmid pF8CIS containing SV40 enhancer and  
DE promoter, cytomegalovirus splice donor site and Ig intron  
KW plasmid pF8CIS; cytomegalovirus; SV40; enhancer; promoter;  
KW stabilising sequence; splice donor intron sequence; Ig region;  
KW splice acceptor sequence; human embryonic kidney cells (293); JM2.  
FH Cytomegalovirus; SV40; +  
FT Key  
FT Location/Qualifiers  
FT Promoter  
FT 1..360  
FT /tag= a  
FT /note="SV40 enhancer and promoter"  
FT 361..580  
FT /tag= b  
FT /note="stabilising sequence which includes  
FT cytomegalovirus donor and intron sequence, the Ig  
FT variable region intron and splice acceptor sequence"  
FT 480..580  
FT /tag= c  
FT /note="IG synthetic 100mer"  
FT 573..590  
FT /tag= d  
FT /note="PstI-ClaI converter"  
FT misc\_rna  
FT 621

FT EP-309235-A.  
PN 29-MAR-1989.  
PD 22-SEP-1988; 308784.  
PF 25-SEP-1987; US-101712.  
PR (GUTH) Genentech Inc.  
PI Gorman CM;  
DR WPI; 89-095394/13.  
PT Transient expression system for recombinant proteins -  
PT encoding trans-activating host cell transfected with vector  
PS Disclousure; 44 pp; English.  
CC Expression vector pF8CIS contains the SV40 enhancer and promoter, the  
CC intron and splice acceptor site, the cDNA encoding factor VIII and the  
CC SV40 polyadenylation and transcription termination sites.  
CC It is used as a method for prodn. of a  
CC desired heterologous protein in a eukaryotic host cell.  
CC The method comprises: transfecting a eukaryotic host cell with a vector  
CC encoding trans-activating protein; transfecting the host cell with  
CC an expression vector having a sequence of a ds DNA comprising, a  
CC stabilising sequence downstream of a promoter and upstream of a DNA  
CC encoding the amino acid sequence of the desired protein, and a poly-  
CC -adenylation sequence upstream to a transcription termination site.  
CC The pref. host cells are human embryonic kidney cells (293) or JM2.  
CC The method provides useful quantities of a desired protein in a  
CC prodn. The transient expression system optimises the interaction between  
CC specific vector components and certain trans-activating proteins.  
SQ Sequence 595 BP; 136 A; 172 C; 146 G; 141 T;

Query Match  
100.0%; Score 10; DB 1; Length 595;

Best Local Similarity 100.0%; Pred. No. 1.1e+03; Indels 0; Gaps 0;  
Matches 10; Conservative 0; Mismatches 0;

OY 1 CTCTCCACAG 10  
DB 543 CTCTCCACAG 552

## RESULT 8

N91649 standard; DNA: 527 BP.  
AC N91649:  
DT 10-MAR-1993 (revised)  
DE 14-MAR-1990 (first entry)  
KW Reshaped light chain variable domain HuVLVLS  
Eukaryotic host.  
PN EP-338745-A.  
PD 25-OCT-1989.  
PF 17-APR-1988; 303767.  
16-APR-1988; GB-009050.  
(CELL) Celltech Ltd.  
OWens RJ, Winter GP, Riechmann L, Yarranton GT, Bodmer MW;  
DR P-PSDB; P91012.  
PT Method of prodn. of recombinant Fv fragments  
PT - by transformed host cells contg. expression vectors  
PS Disclosure; F19. 2; 12pp; English.  
SQ Sequence 527 BP; 136 A; 156 C; 130 G; 105 T;

Query Match 100.0%; Score 10; DB 1; Length 527;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03; Indels 0; Gaps 0;  
Matches 10; Conservative 0; Mismatches 0;

OY 1 CTCTCCACAG 10  
DB 147 CTCTCCACAG 156

## RESULT 9

N91481 standard; DNA: 150 BP.  
AC N91481:  
DT 28-FEB-1990 (first entry)  
DE Start of coding region aequorin/NP fusion protein  
KW Human immunodeficiency virus.  
OS  
WO8909393-A.  
05-OCT-1989.  
20-MAR-1989; U01152.  
24-MAR-1988; US-173231.  
PA (IGEN) Igen Inc.  
PI Kenten JH, Casadei J, Well MJ;  
DR WPI: 89-309634/42.  
PT New luminiscent chimeric proteins - useful in highly sensitive  
PT Immunossays, eg for HIV  
PS Disclosure; page 35; 79pp; English.  
CC used to express the chimeric variable 4-hydroxy-3-nitrophenyl (NP)  
antigen/aequorin protein, in J5581 myeloma or other cells expressing  
CC lambda1 light chain. This protein is used as a diagnostic.  
SQ Sequence 150 BP; 35 A; 38 C; 38 G; 39 T;

Query Match 100.0%; Score 10; DB 1; Length 150;  
Best Local Similarity 100.0%; Pred. No. 1e+03; Indels 0; Gaps 0;  
Matches 10; Conservative 0; Mismatches 0;

OY 1 CTCTCCACAG 10  
DB 118 CTCTCCACAG 127

## RESULT 10

N91645 standard; DNA: 540 BP.

AC N91645:  
DT 15-MAR-1990 (first entry)  
DE Heavy chain of monoclonal antibody 6A4.  
KW Monoclonal antibody 6A4; heavy chain; Pseudomonas aeruginosa; OMP-1.  
FH Key Location/Qualifiers  
FT cds 64..540  
/\*tag= a

EP-338395-A.  
PD 25-OCT-1989.  
PF 12-APR-1989; 106463.  
PR 19-APR-1988; DE-813023.  
PA (BEHM) Behringwerke.  
PI Domdey H, Marger M, von Specht BU;  
DR WPI: 89-310861/43.  
DR P-PSDB; P93079.  
PT Monoclonal antibody to Pseudomonas aeruginosa - and DNA coding for  
PT variable antibody regions.  
PS Claim 1; page 6; 7pp; german.  
CC The sequence has a variable and constant region. Monoclonal antibody 6A4  
reacts with the OMP-1 protein of all 19 known serotypes of P.aeruginosa.  
CC It is used for therapy and diagnosis of infection, and as a carrier for  
CC drugs. The antibody is IgG2a subclass.  
SQ Sequence 540 BP; 145 A; 144 C; 130 G; 121 T;

Query Match 100.0%; Score 10; DB 1; Length 540;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03; Indels 0; Gaps 0;  
Matches 10; Conservative 0; Mismatches 0;

OY 1 CTCTCCACAG 10  
DB 32 CTCTCCACAG 41

## RESULT 11

N91820 standard; DNA: 443 BP.  
AC N91820:  
DT 22-MAR-1990 (first entry)  
DE DNA sequence of the V and J regions of the anti-Tac heavy chain.  
KW Anti-Tac monoclonal antibody; heavy chain; V region; J region;  
KW Interleukin-2 receptor; human T-cells; chimeric antibody.  
OS Mus musculus.  
FH Key Location/Qualifiers  
FT cds 38..442  
/\*tag= a  
WO8909622-A.  
PN 19-OCT-1989.  
PD 13-APR-1989; U01578.  
PF 15-APR-1988; US-181862.  
PA (PROT-) Protein Design Labs, Inc.  
PI Queen CL;  
DR WPI: 89-324085/44.  
DR P-PSDB; P93151.  
PT New interleukin-2 receptor-specific chimeric antibodies - useful for  
PT treating T-cell mediated disorders.  
PS Disclosure; F19. 2; 46pp; English.  
CC The anti-Tac heavy chain is produced by ATCC CRL 9688. It is used in a  
CC mouse/human chimeric antibody, which is specifically reactive with  
CC interleukin-2 receptors on human T-cells.  
SQ Sequence 443 BP; 117 A; 114 C; 110 G; 102 T;

Query Match 100.0%; Score 10; DB 1; Length 443;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03; Indels 0; Gaps 0;  
Matches 10; Conservative 0; Mismatches 0;

OY 1 CTCTCCACAG 10  
DB 6 CTCTCCACAG 15

```
RESULT 12
N91463
ID N91463 standard; DNA; 616 BP.
AC N91463;
DE 24-NOV-1989 (first entry)
DE Sequence of humanised light chain variable domain.
KM Heavy chain variable domain; complementary determining region; antibody;
KM Campath-1; lymphoma; immunosuppression.
OS Rattus norvegicus
FH Key
FH Location/Qualifiers
FT cds
FT 110..154
FT /*tag= a
FT 237..571
FT /*tag= b
PN EP-328404-A.
PD 16-OCT-1989.
PE 10-FEB-1989; 301291.
PF 12-FEB-1988; GB-8803228.
PI (MED) Medical Research Council.
PI Clark MK, Riechmann L, Waldmann H, Winter GP;
DR P-PSDB; P90917.
PT Antibody which binds to antigen Campath-1 - having complementary
PT determining region of rat origin and foreign variable domain framework
PS Disclousure; fig 4; 30pp; English.
CC The sequence encodes rat immunoglobulin heavy chain (see P90917).
CC Misc features a and b show where RNA begins, h and l encode signal
CC peptides, and j, k and l encode complementary determining regions 1, 2
CC and 3 resp.
SQ Sequence 616 BP; 169 A; 174 C; 142 G; 131 T;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 616;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCCTCCACAG 10
Db 228 CTCCTCCACAG 237

RESULT 13
N93069
N93069 standard; DNA; 600 BP.
DE 29-JUN-1990 (first entry)
DE Fragment of pcr.eik.2
KM Tissue plasminogen activator; EIK variant; T7 promoter; Tcs; ds.
FH Key
FH Location/Qualifiers
FT cds
FT 444..600
FT /*tag= a
FT /label= 5' terminal of human t-PA EIK variant gene.
PN MO8810307-A.
PD 29-DEC-1988.
PE 23-JUN-1988; U02127.
PF 24-JUN-1987; US-065794.
PI (GETH) Genetech.
PI Miller HT.
DR WPI; 89-023849/03.
DR P-PSDB; P94866.
PT Recombinant protein mfr. - using balanced inducible transcriptional
PT control system, esp. for mfg. tissue plasminogen activator.
PS Disclousure; fig. 3; 48pp; English.
CC The EIK variant of t-PA is described in EP-199574 and is characterised by
CC substitutional mutations that eliminate a proteolysis site. Plasmid
CC pcr.eik.2 contains CMV enhancer/promoter region, the Ig splice acceptor,
CC the 5p promoter and RNA start, the T7 promoter, the T7E1 leader RBS, the
CC altered preprot-PA gene and the SV40 polyA site.
SQ Sequence 600 BP; 160 A; 157 C; 142 G; 141 T;
```

```
Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 600;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCCTCCACAG 10
Db 267 CTCCTCCACAG 276

RESULT 14
N90647
ID N90647 standard; DNA; 5518 BP.
AC N90647;
DE 01-AUG-1990 (first entry)
DE Nucleotide sequence of plasmid pBG396 including soluble T4-like (ST4)
DE polypeptide number 12 (ST4#12)
KM HIV soluble T4-like polypeptide 12; immunotherapeutic; prophylactic;
KM plasmid pBG396; diagnostic.
OS Homo sapiens.
PN WO8901940-A.
PD 09-MAR-1989.
PE 01-SEP-1988; U02940.
PF 07-JAN-1988; US-141649, US-094322.
PI (BioJ) Bogen Inc.
PI Fisher RA, Gilbert W, Sato VL, Flavell RA, Maraganore JM, Liu TR;
DR WPI; 89-085519/11.
PT DNA sequences coding for soluble T4-like polypeptide(s) -
PT used in immuno:therapeutic and immunosuppressive compns. and for
PT preventing, treating or detecting AIDS
PS Disclousure; ; 207pp; English.
CC It is the nucleotide sequence of the plasmid pBG396. The sequence
CC was isolated from 2 libraries: a lambda gt10 cDNA library derived from T
CC cell tumour line REX and a lambda gt10 cDNA library derived from
CC peripheral blood lymphocytes (PBL). For screening, a series of chemically
CC synthesised antisense oligonucleotide DNA probes based on the known T4
CC protein sequence was used. The polypeptide encoded is useful in
CC immunotherapeutic, prophylactic and diagnostic compns. It can be used
CC to purify HIV from a sample.
SQ Sequence 5518 BP; 1334 A; 1424 C; 1403 G; 1357 T;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 5518;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCCTCCACAG 10
Db 970 CTCCTCCACAG 979

RESULT 15
N90648
ID N90648 standard; DNA; 5566 BP.
AC N90648;
DE 01-AUG-1990 (first entry)
DE Nucleotide sequence of gene encoding mature T4 sequence of plasmid pBG393
DE including soluble T4-like (ST4) polypeptide number 8 (ST4#8)
KM HIV soluble T4-like polypeptide 8; immunotherapeutic; prophylactic;
KM plasmid pBG393; diagnostic.
OS Homo sapiens.
PN WO8901940-A.
PD 09-MAR-1989.
PE 01-SEP-1988; U02940.
PF 07-JAN-1988; US-141649, US-094322.
PI (BioJ) Bogen Inc.
PI Fisher RA, Gilbert W, Sato VL, Flavell RA, Maraganore JM, Liu TR;
DR WPI; 89-085519/11.
PT DNA sequences coding for soluble T4-like polypeptide(s) -
PT used in immuno:therapeutic and immunosuppressive compns. and for
PT preventing, treating or detecting AIDS
PS Disclousure; ; 207pp; English.
CC It is the nucleotide sequence of region encoding the first 182 AA of the
CC mature T4 sequence of plasmid pBG393. It does not contain the additional
```

CC non-T4 lamino acids at the C-terminus following AA 182. The sequence  
CC was isolated from 2 libraries: a lamda g10 cDNA library derived from T  
CC cell tumour line REX and a lamda g10 cDNA library derived from  
CC peripheral blood lymphocytes (PBL). For screening, a series of chemically  
CC synthesised antisense oligonucleotide DNA probes based on the known T4  
CC protein sequence was used. The polypeptide encoded is useful in  
CC immunotherapeutic, prophylactic and diagnostic compns. It can be used  
CC to purify HIV from a sample.  
CC Sequence 5566 BP, 1350 A, 1437 C, 1413 G, 1365 T, 1 Others;

Query Match	100.0%	Score 10	DB 1	Length 5566
Best Local Similarity	100.0%	Pred. No.	1	1e+03
Matches 10	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY	1	CTCTCCACAG	10
Db	970	CTCTCCACAG	979

```

,rich completed: September 18, 1999, 00:33:38
Job time: 18978 sec

```

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:05 ; Search time 2825.05 Seconds  
(without alignments)  
6.982 Million cell updates/sec

Title: US-08-956-518a-92  
Perfect score: 10  
Sequence: 1 CHTCCACAG 10

Scoring table: IDENTITY\_NDC

Searched: 2546578 seqs, 98626752 residues

Database :

EST.\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	10	100.0	437	5	AF119567	AF119567 Homo sapi
2	10	100.0	367	20	D15789	D15789 RICC1259A R
3	10	100.0	270	20	D21568	D21568 MUS69C02 mo
4	10	100.0	436	20	D27930	D27930 CELK06B9R
5	10	100.0	253	20	D28603	D28603 MOS04G11 mo
6	10	100.0	290	20	D31189	D31189 HUML12839 H
7	10	100.0	300	20	D31275	D31275 HUML13208 H
8	10	100.0	233	20	D31359	D31359 HUML13517 H
9	10	100.0	274	20	F03188	F03188 HSC1PA122 n
10	10	100.0	239	20	L26703	L26703 MUSF139A 1a
11	10	100.0	387	20	M77875	M77875 EST01459 Fe
12	10	100.0	335	20	M79545	M79545 WEST00082 M
13	10	100.0	472	20	M79810	M79810 WEST00347 M
14	10	100.0	209	20	M91496	M91496 HUMRPG6AO
15	10	100.0	353	20	T01961	T01961 WEST02682 E
16	10	100.0	400	20	T04427	T04427 474 Lambda-
17	10	100.0	466	20	T04688	T04688 735 Lambda-
18	10	100.0	174	20	T05801	T05801 EST03690 Fe
19	10	100.0	239	20	T07125	T07125 EST05014 Fe
20	10	100.0	362	20	T07194	T07194 EST05083 Fe
21	10	100.0	403	20	T08363	T08363 EST06254 In
22	10	100.0	361	20	T08639	T08639 EST06531 In
23	10	100.0	325	20	T09477	T09477 EST07370 In
24	10	100.0	399	20	T10159	T10159 seq1336 b4H
25	10	100.0	399	20	T10223	T10223 seq1333 b4H
26	10	100.0	252	20	T11222	T11222 hbc949 Huma
27	10	100.0	348	20	T11972	T11972 A843R Heart
28	10	100.0	228	20	T12189	T12189 A564F Heart
29	10	100.0	503	20	T14998	T14998 crs512 Lamb
30	10	100.0	431	20	T15177	T15177 crs661 Lamb
31	10	100.0	250	20	T16196	T16196 IB3703 Infa
32	10	100.0	370	20	T16337	T16337 NIB1145 NOR
33	10	100.0	403	20	T16939	T16939 NIB203 NOR
34	10	100.0	405	20	T18946	T18946 f11040f Tes
35	10	100.0	214	20	T23678	T23678 seq315 N-b4
36	10	100.0	259	20	T23964	T23964 seq2117 2HB
37	10	100.0	269	20	T24043	T24043 seq2201 2HB
38	10	100.0	450	20	T27084	T27084 NIBT261G03R
39	10	100.0	353	20	T27119	T27119 NIBT300E12R
40	10	100.0	480	20	T27137	T27137 NIBT314E12R
41	10	100.0	461	20	T27397	T27397 hbc2545 Hum
42	10	100.0	489	20	T28131	T28131 EST29787 Hu
43	10	100.0	382	20	T28546	T28546 EST47957 Hu
44	10	100.0	382	20	T29034	T29034 EST65948 Hu
45	10	100.0	211	54	HSM009833	A1044983 Homo sapi

#### ALIGNMENTS

RESULT 1  
ID AF119567/c standard; RNA; EST; 437 BP.  
XX AC AF119567;  
XX SV AF119567.1  
XX

NI 94325172  
 XX 11-MAR-1999 (Rel. 59, Created)  
 DT 11-MAR-1999 (Rel. 59, Last updated, Version 1)  
 XX  
 DE Homo sapiens clone IEDP1-2, mRNA sequence.  
 XX  
 XX EST.  
 XX  
 OS Homo sapiens (human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;  
 CC Primates; Catarrhini; Homidae; Homo.  
 XX  
 XX [1]  
 XX 1-437  
 RP MEDLINE: 97471696.  
 RX Jacob A.N., Baskaran N., Kandpal G., Narayan D., Bhargava A.K.,  
 RA Kandpal R.P.;  
 RT "Isolation of human ear specific cDNAs and construction of cDNA  
 libraries from surgically removed small amounts of inner ear tissues";  
 Somat. Cell Mol. Genet. 23(2):83-95(1997).  
 --  
 RN [2]  
 RP 1-437  
 RA Nellisery J.K., Kandpal R.P.;  
 RT Submitted (13-JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RL Fels Institute for Cancer Research and Molecular Biology, Temple  
 RL University School of Medicine, 3307 North Broad Street, Philadelphia, PA  
 RL 19140, USA  
 XX  
 XX Key Location/Qualifiers  
 FH 1. 437  
 FT source  
 FT /db\_xref="taxon:9606"  
 FT /note="Isolated by kinetic enrichment in a representational  
 difference analysis (RDA) experiment; similar to sequence  
 encoding MHC class II HLA-RP1"  
 FT /organism="Homo sapiens"  
 FT /tissue\_type="inner and middle ear"  
 FT /dev\_stage="fetus"  
 FT /clone="IEDP1-2"  
 XX  
 SQ Sequence 437 BP; 114 A; 93 C; 139 G; 77 T; 14 other;  
 Query Match 100.0%; Score 10; DB 5; Length 437;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 CTCCTCCACAG 10  
 Db 168 CTCCTCCACAG 159  
 RESULT 2  
 D15789/c 367 bp mRNA EST 20-JUL-1998  
 LOCUS D15789  
 DEFINITION RIC1259A Rice callus Oryza sativa cDNA clone R, mRNA sequence.  
 ACCESSION D15789  
 NID 9286983  
 VERSION D15789.1 GI:286983  
 KEYWORDS EST.  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
 Poaceae; Oryza.  
 1 (bases 1 to 367)  
 Sasaki, T. and Minobe, Y.  
 Rice cDNA from callus  
 Unpublished (1994)  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Contact: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program  
 2-1-2 Kannondai, Tsukuba  
 Ibaraki,  
 Japan 305  
 Tel: 0298-38-7441  
 Fax: 0298-38-7468  
 Email: tsasaki@agr.affrc.go.jp  
 PROJECT "RGP"  
 Seq primer: oligo(dT).  
 Location/Qualifiers  
 1. 367  
 /organism="Oryza sativa"  
 /strain="cultivar Nipponbare, sub-species Japonica"  
 /db\_xref="taxon:4530"  
 /clone="R"  
 /note="Vector: pBluescript II SK+; Site\_1: SalI; Site\_2:  
 NotI; cDNA prepared from rice callus RNAs by using  
 oligo(dT) as a primer and ligating to the SalI-NotI site  
 of pBluescript II SK+ phagemid."  
 BASE COUNT 53 a 114 c 127 g 69 t 4 others  
 ORIGIN  
 Query Match 100.0%; Score 10; DB 20; Length 367;  
 Best Local Similarity 100.0%; Pred. No. 8.9e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 CTCCTCCACAG 10  
 Db 125 CTCCTCCACAG 116  
 RESULT 3  
 D21568 270 bp mRNA EST 07-OCT-1996  
 LOCUS D21568  
 DEFINITION MOS69C02 mouse embryonal carcinoma cell line F9 Mus musculus cDNA  
 clone 69C02, mRNA sequence.  
 ACCESSION D21568  
 NID 6618685  
 VERSION D21568.1 GI:618685  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 270)  
 Nishiguchi, S., Joh, T., Horie, K., Zou, Z., Yasunaga, T. and Shimada, K.  
 A survey of genes expressed in undifferentiated mouse embryonal  
 carcinoma F9 cells: characterization of low-abundance mRNAs  
 J. Biochem. 116, 128-139 (1994)  
 JOURNAL  
 MEDLINE  
 COMMENT  
 Contact: Kazunori Shimada  
 Department of Medical Genetics, Division of Molecular Biomedicine  
 Research Institute for Microbial Diseases, Osaka University  
 3-1, Yamadaoka, Suita, Osaka, 565, Japan  
 Tel: 06-879-8325  
 Fax: 06-879-8326  
 FEATURES  
 SOURCE 1. 270  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="69C02"  
 /clone\_1b="mouse embryonal carcinoma cell line F9"  
 BASE COUNT 96 a 63 c 81 g 26 t 4 others  
 ORIGIN  
 Query Match 100.0%; Score 10; DB 20; Length 270;  
 Best Local Similarity 100.0%; Pred. No. 8.4e+03;



Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10  
|||||  
Db 99 CTCCTCCACAG 108

RESULT 4  
D27930/c 436 bp mRNA EST 20-NOV-1995  
LOCUS CELK00689R Yuj1 Kohara unpublished cDNA Caenorhabditis elegans cDNA  
DEFINITION clone ykcb9 3', mRNA sequence.  
D27930 6522671  
ACCESSION  
NID D27930.1 GI:522671  
VERSION  
KEYWORDS EST.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
Rhabditina; Rhabditioidea; Rhabditidae; Peloderiinae; Caenorhabditis.  
1 (bases 1 to 436)  
Kohara, T., Mitsuki, H., Nishigaki, A., Mochizuki, T., Sugimoto, A. and  
Tabara, H.  
Toward an expression map of the C. elegans genome  
Unpublished (1994)

TITLE JOURNAL  
COMMENT

CONTACT: Yuj1 Kohara  
Gene Library Lab  
National Institute of Genetics  
Yata 1111, Mishima, Shizuoka 411, Japan  
Tel: 0559-75-0771  
Fax: 0559-75-6240  
Email: ykoha@nig.nig.ac.jp.  
Location/Qualifiers

FEATURES  
Source  
1. 436  
/organism="Caenorhabditis elegans"  
/strain="CB1489 him-8(el489)"  
/db\_xref="taxon:6239"  
/map="18 q21.1 q21.3"  
/clone="ykcb9"  
/clone\_lib="Yuj1 Kohara unpublished cDNA"  
/note="dev-stage=varied, sex=Hermaphrodite male,  
tissue-type=whole animal"

BASE COUNT 140 a 89 c 82 g 125 t

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 436;  
Best Local Similarity 100.0%; Pred. No. 9.3e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10  
|||||  
Db 49 CTCCTCCACAG 40

RESULT 5  
D28603 253 bp mRNA EST 07-OCT-1996  
LOCUS M0504G11 mouse embryonal carcinoma cell line F9 Mus musculus cDNA  
DEFINITION clone 04G11, mRNA sequence.  
D28603 6618497  
ACCESSION  
NID D28603.1 GI:618497  
VERSION  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 253)  
Nishiguchi, S., Joh, T., Horie, K., Zou, Z., Yasunaga, T. and Shimada, K.  
A survey of genes expressed in undifferentiated mouse embryonal  
carcinoma F9 cells: characterization of low-abundance mRNAs

JOURNAL T. Biochem. 116, 128-139 (1994)  
MEDLINE 95096008  
COMMENT

CONTACT: Kazunori Shimada  
Department of Medical Genetics, Division of Molecular Biomedicine  
Research Institute for Microbial Diseases, Osaka University  
3-1, Yamadaoka, Suita, Osaka, 565, Japan  
Tel: 06-879-8325  
Fax: 06-879-8326.  
Location/Qualifiers

FEATURES  
Source  
1. 253  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="04G11"  
/clone\_lib="mouse embryonal carcinoma cell line F9"

BASE COUNT 56 a 82 c 58 g 50 t 7 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 253;  
Best Local Similarity 100.0%; Pred. No. 8.3e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10  
|||||  
Db 101 CTCCTCCACAG 110

RESULT 6  
D31189 290 bp mRNA EST 08-FEB-1995  
LOCUS HUM12839 Human fetal lung Homo sapiens cDNA 5', mRNA sequence.  
DEFINITION  
D31189 9644069  
ACCESSION  
NID  
VERSION D31189.1 GI:644069  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 290)  
Sudo, K., Chinen, K. and Nakamura, Y.  
2038 expressed sequence tags (ESTs) from a human fetal lung cDNA  
library  
Genomics 24, 276-279 (1995)  
95213017

JOURNAL  
MEDLINE  
COMMENT

CONTACT: Yusuke Nakamura  
Institute of Medical Science  
University of Tokyo  
4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan  
Tel: 81-3-5449-5372  
Fax: 81-3-5449-5433  
Email: yusuke@ims.u-tokyo.ac.jp  
Insert Length: 825 Std Error: 0.00  
High quality sequence stop: 1.  
Location/Qualifiers

FEATURES  
Source  
1. 290  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Human fetal lung"

BASE COUNT 66 a 78 c 78 g 61 t 7 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 290;  
Best Local Similarity 100.0%; Pred. No. 8.3e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10  
|||||  
Db 126 CTCCTCCACAG 135

```

RESULT 7
LOCUS D31275 300 bp mRNA EST 08-FEB-1995
DEFINITION HUM13208 Human fetal lung Homo sapiens CDNA 5', mRNA sequence.
ACCESSION D31275
NID 9644155
VERSION D31275.1 GI:644155
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 300)
AUTHORS Sudo,K., Chinen,K. and Nakamura,Y.
TITLE 2058 expressed sequence tags (ESTs) from a human fetal lung cDNA
JOURNAL GENOMICS 24, 276-279 (1995)
MEDLINE 95213017
ENT

Contact: Yusuke Nakamura
Institute of Medical Science
University of Tokyo
4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan
Tel: 81-3-5449-5372
Fax: 81-3-5449-5433
Email: yusuke@ims.u-tokyo.ac.jp
Insert Length: 2137 Std Error: 0.00
High quality sequence stop: 377.
Location/Qualifiers
1. 300
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="710B10; 15"
/clone_lib="Human fetal lung"
BASE COUNT 70 a 83 c 82 g 63 t 2 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 300;
Best Local Similarity 100.0%; Pred. No. 8.6e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10
Db 128 CTCCTCCACAG 137

...SULT 8
LOCUS D31359 233 bp mRNA EST 08-FEB-1995
DEFINITION HUM13117 Human fetal lung Homo sapiens CDNA 5', mRNA sequence.
ACCESSION D31359
NID 9644239
VERSION D31359.1 GI:644239
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 233)
AUTHORS Sudo,K., Chinen,K. and Nakamura,Y.
TITLE 2058 expressed sequence tags (ESTs) from a human fetal lung cDNA
JOURNAL GENOMICS 24, 276-279 (1995)
MEDLINE 95213017
ENT

Contact: Yusuke Nakamura
Institute of Medical Science
University of Tokyo
4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan
Tel: 81-3-5449-5372
Fax: 81-3-5449-5433

```

```

Email: yusuke@ims.u-tokyo.ac.jp
Insert Length: 586 Std Error: 0.00
High quality sequence stop: 400.
Location/Qualifiers
1. 233
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human fetal lung"
BASE COUNT 45 a 72 c 59 g 55 t 2 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 233;
Best Local Similarity 100.0%; Pred. No. 8.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10
Db 192 CTCCTCCACAG 201

RESULT 9
LOCUS F03188/c 274 bp mRNA EST 02-FEB-1995
DEFINITION HSC1PA122 normalized infant brain cDNA Homo sapiens cDNA clone
c-1pa12 3', mRNA sequence.
ACCESSION F03188
NID 9646745
VERSION F03188.1 GI:646745
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 274)
AUTHORS Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
Lorenzo,F., Mitchell,H., Mariage-Sanson,R., Pietu,G., Pouillot,Y.,
Sebastiani-Kabaktchis,C. and Tessier,A.
IMAGE: molecular integration of the analysis of the human genome
and its expression
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
95277534
JOURNAL
MEDLINE
COMMENT

Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read. removed at sequence 5'end
Genexpress_library_idt: C; Genexpress_sequence_idt: alc-1pa12
Seq primer: (-21)M13 universal.
Location/Qualifiers
1. 274
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="X g24-qter"
/clone="c-1pa12"
/sex="Female"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/notes="Organ: brain; Vector: lafmid BA; Site.1: HindIII;
Site.2: NotI; sex:Female; dev:stage=3 months old;
isolate-muscular atrophy patient; tissue-type=total
brain; total mRNA was oligo (dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
laifmid BA vector. Clone library from B.Souares, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"
BASE COUNT 65 a 70 c 53 g 86 t

```

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 274;  
 Best Local Similarity 100.0%; Pred. No. 8.4e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10  
 |||||

Db 180 CTCCTCCACAG 171

RESULT 10  
 L26703 239 bp mRNA EST 01-MAR-1995  
 LOCUS L26703/c  
 DEFINITION MMSF139A lambda unizap male testis Mus musculus cDNA clone F139,  
 mRNA sequence.  
 ACCESSION L26703  
 "ID 9437492  
 RSTON L26703.1 GI:437492  
 EST.  
 "WORDS house mouse;  
 SOURCE house mouse;  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 239)  
 AUTHORS Kerr,S.M., Vambrie,S., McKay,S.J. and Cooke,H.J.  
 TITLE Analysis of cDNA sequences from mouse testis  
 JOURNAL Mamm. Genome 5, 557-565 (1994)  
 MEDLINE 95093181  
 COMMENT Contact: Kerr,S.M., Vambrie,S., McKay,S.J. and Cooke,H.J.  
 Location/Qualifiers  
 1..239  
 /organism="Mus musculus"  
 /strain="Swiss"  
 /db\_xref="taxon:10090"  
 /clone="F139"  
 /note="Vector: lambda unizap male testis"  
 /note="Vector: lambda unizap; lambda unizap library from  
 mail mouse testis"

BASE COUNT 66 a 62 c 78 g 32 t 1 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 239;  
 Best Local Similarity 100.0%; Pred. No. 8.2e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10  
 |||||

Db 180 CTCCTCCACAG 171

RESULT 11  
 M77875 387 bp mRNA EST 26-MAY-1992  
 LOCUS M77875  
 DEFINITION EST01459 Fetal brain, strata gene (cat#936206) Homo sapiens cDNA  
 clone HFBCA49 similar to Ribosomal protein Y110, mRNA sequence.  
 ACCESSION M77875  
 "ID 9273613  
 RSTON M77875.1 GI:273613  
 EST.  
 "WORDS human;  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 387)  
 AUTHORS Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,  
 Utterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.  
 TITLE Sequence identification of 2,375 human brain genes  
 JOURNAL Nature 355, 632-634 (1992)  
 MEDLINE 92168112

COMMENT

Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org  
 Seq primer: M13 Forward.  
 Location/Qualifiers  
 1..387  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (host):77646"  
 /db\_xref="GDB:D0885E"  
 /db\_xref="taxon:9606"  
 /clone="HFBCA49"  
 /note="Vector: Fetal brain, strata gene (cat#936206)"  
 /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;  
 oligo-dt + random primed cDNA synthesis; lambdaZAP-II  
 vector, 1.0kb average insert size."

BASE COUNT 87 a 100 c 104 g 94 t 2 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 387;  
 Best Local Similarity 100.0%; Pred. No. 9e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10  
 |||||

Db 96 CTCCTCCACAG 105

RESULT 12  
 M79545 335 bp mRNA EST 30-JUN-1992  
 LOCUS M79545/c  
 DEFINITION WEST00082 Mixed stage, Strata gene (cat. #937006) Caenorhabditis  
 elegans cDNA clone CEMS452 similar to 3-Methyl-2-oxobutanoate  
 dehydrogenase homologous peptide, mRNA sequence.  
 ACCESSION M79545  
 "ID 9271570  
 RSTON M79545.1 GI:271570  
 EST.  
 "WORDS Caenorhabditis elegans.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans  
 Eukaryota; Metazoa; Nematoda; Secernentes; Rhabdita; Rhabditida;  
 Rhabditina; Rhabditidae; Rhabditidae; Pelodermidae; Caenorhabditis.  
 REFERENCE 1 (bases 1 to 335)  
 AUTHORS McCombie,W.R., Adams,M.D., Kelley,J.M., Fitzgerald,M.G.,  
 Utterback,T.R., Khan,M., Dubnick,M., Kerlavage,A.R., Venter,J. and  
 Fields,C.  
 TITLE Caenorhabditis elegans expressed sequence tags identify gene  
 families and potential disease gene homologues  
 JOURNAL Nature Genet. 1, 124-131 (1992)  
 MEDLINE 93250983  
 COMMENT Other ESTs: WEST00081  
 Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org  
 Seq primer: M13 Reverse.  
 Location/Qualifiers  
 1..335  
 /organism="Caenorhabditis elegans"  
 /db\_xref="taxon:6239"  
 /clone="CEMS452"  
 /clone.lib="Mixed stage, Strata gene (cat. #937006)"  
 /note="Vector: Uni-ZAP XR; C. elegans mixed stage cDNA  
 library. Strata gene catalog #937006. The library is oligo  
 dt primed and directionally cloned in the Uni-ZAP XR

BASE COUNT 96 a 64 c 88 g 80 t 7 others  
 ORIGIN vector."

Query Match 100.0%; Score 10; DB 20; Length 335;  
 Best Local Similarity 100.0%; Pred. No. 8.8e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10  
 |||||

Db 225 CTCCTCCACAG 216

RESULT 13  
 M79810/c 472 bp mRNA EST 30-JUN-1992  
 LOCUS W5700347 Mixed stage, Strata gene (cat. #937006) Caenorhabditis  
 DEFINITION elegans cDNA clone CEMSE18, mRNA sequence.  
 SSION M79810.  
 M79810. 9271829  
 M79810.1 GI:271829  
 EST.  
 KEYWORDS Caenorhabditis elegans.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
 Rhabditina; Rhabditioidea; Rhabditidae; Pelodierinae; Caenorhabditis.  
 REFERENCE 1 (bases 1 to 472)  
 AUTHORS McCombie, W.R., Adams, M.D., Kelley, J.M., Fitzgerald, M.G.,  
 Uterback, T.R., Khan, M., Dubnick, M., Kerlavage, A.R., Venter, J. and  
 Fields, C.  
 TITLE Caenorhabditis elegans expressed sequence tags identify gene  
 families and potential disease gene homologues  
 JOURNAL Nature Genet. 1, 124-131 (1992)  
 MEDLINE 93250983  
 COMMENT Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@ligr.org  
 Seq primer: M13 Forward.  
 Location/Qualifiers  
 1. 472  
 /db\_xref="taxon:6239"  
 /organism="Caenorhabditis elegans"  
 /clone\_lib="Mixed stage, Strata gene (cat. #937006)"  
 /note="Vector: Uni-ZAP XR; C. elegans mixed stage cDNA  
 library. Strata gene catalog #937006. The library is oligo  
 dt primed and directionally cloned in the Uni-ZAP XR  
 vector."

BASE COUNT 123 a 112 c 95 g 141 t 1 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 472;  
 Best Local Similarity 100.0%; Pred. No. 9.4e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10  
 |||||

Db 444 CTCCTCCACAG 435

RESULT 14  
 M91496 209 bp mRNA EST 29-OCT-1992  
 LOCUS HUMR1PGE80 Subtracted human retinal pigment epithelium (RPE) Homo  
 DEFINITION sapiens cDNA, mRNA sequence.  
 ACCESSION M91496

NID 9275912  
 VERSION M91496.1 GI:275912  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 209)  
 AUTHORS Gieser, L. and Swaroop, A.  
 TITLE Expressed sequence tags and chromosomal localization of cDNA clones  
 from a subtracted retinal pigment epithelium library  
 JOURNAL Genomics 13, 873-876 (1992)  
 MEDLINE 92347897  
 COMMENT Contact: Swaroop, A.  
 Department of Ophthalmology  
 Kellogg Eye Center, University of Michigan  
 1000 Wall Street, Ann Arbor, MI 48105.  
 Location/Qualifiers  
 1. 209  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Subtracted human retinal pigment epithelium  
 (RPE) library"

BASE COUNT 71 a 48 c 38 g 50 t 2 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 209;  
 Best Local Similarity 100.0%; Pred. No. 7.9e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10  
 |||||

Db 97 CTCCTCCACAG 106

RESULT 15  
 T01961 353 bp mRNA EST 10-NOV-1992  
 LOCUS W5702682 Early embryo, Strata gene (cat. #937007) Caenorhabditis  
 DEFINITION elegans cDNA clone CEMSx81, mRNA sequence.  
 ACCESSION T01961  
 T01961.1 GI:278442  
 EST.  
 KEYWORDS Caenorhabditis elegans.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
 Rhabditina; Rhabditioidea; Rhabditidae; Pelodierinae; Caenorhabditis.  
 REFERENCE 1 (bases 1 to 353)  
 AUTHORS McCombie, W.R., Kelley, J.M., Aubin, L., Goscochea, M.,  
 Fitzgerald, M.G., Wu, A., Adams, M.D., Dubnick, M., Kerlavage, A.R.,  
 Venter, J.C. and Fields, C.A.  
 TITLE Caenorhabditis elegans CDNA  
 JOURNAL Unpublished (1993)  
 COMMENT Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@ligr.org  
 Seq primer: M13 Forward.  
 Location/Qualifiers  
 1. 353  
 /organism="Caenorhabditis elegans"  
 /db\_xref="taxon:6239"  
 /clone\_lib="Early embryo, Strata gene (cat. #937007)"

BASE COUNT 124 a 38 c 87 g 99 t 5 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 353;  
Best Local Similarity 100.0%; Pred. No. 8.9e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Caps 0;  
QY 1 CTCCTCACAG 10  
|||||||  
Db 279 CTCCTCACAG 270

Search completed: September 17, 1999, 21:28:06  
Job time: 14270 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:33:38 ; Search time 425.19 Seconds  
(without alignments)  
5.884 Million cell updates/sec

Title: US-08-956-518a-93

Perfect score: 10

Sequence: 1 GTCCTCCGAG 10

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	100.0	4885	003618	Phosphoenolpyruvate
2	10	100.0	1845	N90601	HindIII fragment o
3	10	100.0	829	N90234	Gastrin recombinan
4	10	100.0	2045	N90749	Human colony stimu
5	10	100.0	5406	N93197	Fragment of clone
6	10	100.0	2024	003143	Nucleotide sequenc
7	10	100.0	2743	004799	Recombinant sequen
8	10	100.0	1071	N81654	LVD regions of hu
9	10	100.0	1399	006280	Sequence encoding
10	10	100.0	15464	011415	Ryanodine receptor
11	10	100.0	1196	N91627	Sequence of gamma-
12	10	100.0	14928	011707	Immunoglobulin D-r
13	10	100.0	861	N60588	First half of part
14	10	100.0	549	N60295	DNA encoding hepat
15	10	100.0	4382	012759	P40 genomic DNA. N
16	10	100.0	973	013171	Retinal cGMP phosph
17	10	100.0	2156	013241	Human HSF cDNA seq
18	10	100.0	12364	013547	ACV gene pcDNA. M
19	10	100.0	12364	013607	ACV synthetase gen
20	10	100.0	3014	014954	Human dopamine DI
21	10	100.0	15155	N50107	DNA encoding Facto
22	10	100.0	1515	N50360	Endoglucanase ENO
23	10	100.0	1516	N50179	Sequence of endogl
24	10	100.0	1170	020217	Sequence of tuft3 g
25	10	100.0	4258	022439	DNA of hgc-co-1, en
26	10	100.0	5136	023133	DNA encoding mask1
27	10	100.0	13875	020240	Human centromeric
28	10	100.0	552	N00001	Sequence encoding
29	10	100.0	2743	N00003	Sequence of a part
30	10	100.0	2156	025713	Sequence of human
31	10	100.0	1191	027829	Fetal calf TP-like
32	10	100.0	1282	029958	5-Substid. hydrantio
33	10	100.0	1317	031878	Cyclin D1 promoter
34	10	100.0	3980	036780	Human cardiac cgt
35	10	100.0	2560	035298	Murine TC-CSF. New
36	10	100.0	973	043549	Gammma subunit of h
37	10	100.0	13747	047957	Myotonic dystrophy
38	10	100.0	4099	047957	Human endothelial
39	10	100.0	397	060714	Human brain expres
40	10	100.0	338	060300	Human vitamin D re
41	10	100.0	1399	051424	Human cyclin D1 pr
42	10	100.0	1316	053210	Neurocan DNA. Euka
43	10	100.0	5191	057710	

## ALIGNMENTS

44 10 100.0 1593 1 053404  
45 10 100.0 116277 1 X20249

Eleusine indica L  
Borrelia burgdorferi

```

RESULT 1
ID 003618 standard; DNA: 4885 BP.
AC 003618:
DT 01-AUG-1990 (first entry)
DE Phosphoenolpyruvate carboxylase (ppc) gene inserted into plasmid pPG1200.
OS Phosphoenolpyruvate carboxylase; ppc; pPG1200; ds.
KW Corynebacterium glutamicum.
FH Key location/Qualifiers
FT cds 921..3677
FT /*tag= a

PD EP-358940-A.
PD 21-MAR-1990.
PF 8-AUG-1989; 114632.
PR 12-SEP-1988; GB-021319.
PA (DEBS) Degussa AG.
PI Bachmann B, Thierbach GD, Kalinowski J, Puhler A, O'Reagan M,
PI Viret JF, Lepage P, Lemoine Y;
DR WPI: 90-084832/12.
DR P-PSDB: R05512.
PT Deoxyribonucleic acid fragment, for L-amino acid -
PT obtd. from corynebacterium glutamicum strain coding for
PT phosphoenolpyruvate carboxylase.
PS Disclosure; P: English.
CC PPC is important as a key enzyme in the anaerobic function of the cell,
CC by increasing oxaloacetate levels. It indirectly increases biosynthetic
CC levels of L-IAs.
SQ Sequence 4885 BP; 1071 A; 1356 C; 1318 G; 1139 T; 1 Others;

Query Match 100.0%; Score 10; DB 1; Length 4885;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCGAG 10
DB 4295 GTCCTCCGAG 4286

RESULT 2
ID N90601 standard; DNA: 1845 BP.
AC N90601:
DT 05-JAN-1990 (first entry)
DE HindIII fragment of human CSF-1.
OS Colony stimulating factor; Immunosuppression; lymphokine.
KW Homo sapiens
FH Key location/Qualifiers
FT exon 218..340
FT /*tag= a
FT /*tag= b
FT /*tag= .1449
FT /*tag= 1387
FT /*tag= 821..822
FT /*tag= c

PD US4847201-A.
PD 11-SEP-1989.
PR 09-FEB-1988; 157094.
PR 05-FEB-1985; US-876819.
PA (Cetus) Cetus Corp.
PI Kasawaki ES, Lachner MB, Van Arsdel JN, Wang AM, Ralph P, Coyen MY,
PI WPI: 89-255732/35.
DR P-PEDB: P91458.
PT Human and murine colony stimulating factors, CSF-1 - used to overcome
PT immunosuppression caused by eg chemotherapy, and obtd. by recombinant
PT methods.
PS Disclosure; fig 4a-4c; 30pp; English.

```

CC Feature c represents 200 unnamed bases. Source of mRNA was human  
 CC pancreatic carcinoma cell line MiPaCa-2. See also P91458 and P91459.  
 SQ Sequence 1845 BP; 410 A; 456 C; 513 G; 466 T;

Query Match 100.0%; Score 10; DB 1; Length 1845;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCAG 10  
 Db 525 GTCCTCCAG 534

## RESULT 3

ID N90234 standard; DNA; 829 BP.

AC N90234; 1-NOV-1989 (first entry)

DT Gastrin recombinant DNA.

Recombinant DNA: promoter; signal peptide; gastrin; shuttle  
 vector; yeast; Escherichia coli.

Key Location/Qualifiers

OS Key  
 FH exon 1..472  
 FT /\*tag- a  
 FT 473..602  
 FT /\*tag- b  
 FT 603..829  
 FT /\*tag- c  
 FT 78..81  
 FT /\*tag- d  
 FT 288..291  
 FT /\*tag- e  
 FT 320..325  
 FT /\*tag- f  
 FT 436..441  
 FT /\*tag- g  
 FT 711..716  
 FT /\*tag- h  
 FT 768..771  
 FT /\*tag- i  
 FT 794..797  
 FT /\*tag- j  
 FT 262..694  
 FT /\*tag- k  
 PN J64002579-A.  
 03-MAY-1989.  
 24-JUN-1987; 156745.  
 24-JUN-1987; JP-156745.  
 PA (AJIN) Ajinomoto KK.  
 DR WPI; 89-176418/24.  
 DR P-PSDB; P90331.  
 PT Recombinant DNA encoding gastrin - also comprising promoter and signal  
 PT peptide coding regions. Inserted into shuttle vector to produce  
 PT gastrin in yeast or E.coli.  
 PS Claim: page 524 and fig 2: 10pp; Japanese.  
 CC The recombinant DNA effectively produces gastrin in high quantities.  
 CC It is made by integrating DNA fragment encoding gastrin into a shuttle  
 CC vector which can grow in yeast or E.coli, eg pAM82, CY37, pAT77, YEP51,  
 CC YEP52 or pJDB218. The DNA encoding signal peptide is, eg from human  
 CC gastrin, Mfalpalm (Yeast secretion hormone), yeast invertase, killer  
 CC toxin, yeast alkaline or acid phosphatase, human lysozyme, human alpha-  
 CC amylase or human trypsin inhibitor. DNA encoding gastrin is from, eg  
 CC pHE53. See also P90331. Misc. feature d is Sau3A site, e is Sau3A site,  
 CC f is HindIII site, g is PvuII site, h is HindIII site, i is Sau3A site,  
 CC and j is Sau3A site. The CDS is minus the intron.  
 CC Bases 262-318 encode signal peptide, and bases 319-429 encode gastrin.  
 SQ Sequence 829 BP; 199 A; 251 C; 212 G; 167 T;

Query Match 100.0%; Score 10; DB 1; Length 829;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCAG 10  
 Db 524 GTCCTCCAG 533

## RESULT 4

ID N90749 standard; CDNA; 2045 BP.

AC N90749; 06-JUN-1990 (first entry)

DT Human colony stimulating factor-1 (CSF-1)

DE Human colony stimulating factor-1; CSF-1; cytomagalovirus; CMV;

KW Gram-negative sepsis; sarcoma therapy; melanoma therapy; ADCC;

OS Homo sapiens.

Key

FT Introns Location/Qualifiers  
 FT 1..217  
 FT /\*tag- a  
 FT 218..340  
 FT /\*tag- b  
 FT 341..1581  
 FT /\*tag- c  
 FT 1582..1650  
 FT /\*tag- d  
 PN W08902746-A.  
 PD 06-APR-1989.  
 PF 16-SEP-1988; US-243253, US-099872.  
 PR 14-SEP-1988; US-243253, US-099872.  
 PA (CERV) Cetus Corp.  
 PI Ralph P. Warren MK, Chong KT, Devlin JT, Zimmerman R;  
 DR WPI; 89-114251/15.  
 DR P-PSDB; P93555; P93556.  
 PT Use of recombinant colony stimulating factor-1 -  
 PT for enhancing prodn. of white blood cells or treating viral or  
 PT bacterial infectious diseases or tumours  
 PS Figure 2a-2c; pp. 3/7-5/7; 62pp; English.  
 CC It is the sequenced portion of a 3.9kb HindIII fragment encoding human  
 CC CSF-1 sequences. It is used in the mfr. of a compsn. useful for  
 CC enhancing prodn. of white blood cells, including neutrophils, from stem  
 CC cells. It is also used in mfr. of a compsn. useful for treatment/  
 CC prophylaxis of bacterial or viral infectious diseases (esp.  
 CC cytomegalovirus or Gram-negative sepsis), for the treatment  
 CC of sarcoma or melanoma, for stimulating antibody-dependent cellular  
 CC cytotoxicity (ADCC), and for wound healing.  
 SQ Sequence 2045 BP; 410 A; 455 C; 514 G; 466 T; 200 Others;

Query Match 100.0%; Score 10; DB 1; Length 2045;

Best Local Similarity 100.0%; Pred. No. 1.1e+03;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCAG 10  
 Db 525 GTCCTCCAG 534

## RESULT 5

ID N93197 standard; DNA; 5406 BP.

AC N93197; 28-JUN-1990 (first entry)

DE Fragment of clone lambda HK65a contg. kallikrein gene.

KW Kallikrein gene; vasodilator; male infertility; lambda HK65a; ss.

OS Homo sapiens.

Key

FT exons Location/Qualifiers  
 FT 803..848  
 FT /\*tag- a  
 FT 2538..2697  
 FT /\*tag- b  
 FT 3967..4256  
 FT /\*tag- c

```
FT exon 4375..4334
FT      /*tag= d
FT exon 4838..5075
FT      /*tag= e
PN EP-297913-A.
PD 04-JAN-1989.
PF 30-JUN-1988; 306039.
PR 30-JUN-1987; US-068594.
PA (AMGE-) Amgen Inc.
PI Lin FK, Lu HS;
DR WPI: 89-009139/02.
P-PSDB: P93722.
PT New recombinant kallikrein polypeptide(s) and encoding DNA -
PS useful as vasodilators and for treating male infertility.
PT Table V; page 13; 43pp; English.
CC Lambda HK64a is a subclone produced from clones isolated from a CHA phage
CC bouane human fetal liver genomic library using monkey kallikrein MKK80a
CC clone DNA as a probe.
See also M93193.8.
Sequence 3406 BP; 974 A; 1864 C; 1309 G; 1259 T;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 5406;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCCGAG 10
DB 1223 GTCCTCCCGAG 1232

RESULT 6
ID 003143/c
AC 003143:
DR 10-MAR-1993 (revised)
DI 09-SEP-1990 (first entry)
DE Nucleotide sequence of cDNA from Aspergillus niger encoding glucose
DE oxidase
KW Glucose oxidase; Aspergillus niger; glucose detection;
KW glucose estimation.
OS Aspergillus niger.
FH Key Location/Qualifiers
FT misc_difference 2..2025
FT mat_peptide
/*tag= a
/*note="nucleotide sequence as given in Fig 5B-1-5B-6"
98..1849
/*tag= b
32..97
/*tag= c
/product=peptide
cds
/*tag= a
WO8912675-A.
28-DEC-1989.
PN 20-JUN-1988; U02696.
PF 21-JUN-1988; US-209530.
PR 19-JUN-1989; US-366377.
PA (CHIR-) Chiron Corp.
PI Rosenberg S;
DR WPI: 90-022535/03.
P-PSDB: R04724.
PT Recombinant polynucleotide(s) encoding glucose oxidase or a mutein -
PT and vectors contg. them, secreting glucose oxidase or
PT hyper-glycosylated analogues into medium
PS Disclosure; B-6; 108pp; English.
CC The cdna can also be used to isolate GO-encoding sequences from other
CC sources. Its encoded GO is used in desugaring eggs, removing oxygen from
CC moist food prods., beverages, flavours and hermetically-sealed food
CC packages, and in detection and estimation of glucose in industrial solns.
CC and body fluids, eg blood and urine. This sequence is the same as that
CC given in Fig 5B of the patent, but with the addition of one Guanine at
CC the start of the sequence.
Sequence 2024 BP; 450 A; 615 C; 515 G; 445 T;
```

```
Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 2024;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCCGAG 10
DB 1559 GTCCTCCCGAG 1550

RESULT 7
ID 004799/c
AC 004799:
DR 30-OCT-1990 (first entry)
DE Recombinant sequence encoding hepatitis B antigens.
KW Hepatitis B virus; vaccine; HBV; ds.
OS Synthetic.
FH Key Location/Qualifiers
FT cds
/*tag= a
1..80
FT cds
/*tag= b
87..634
FT cds
/*tag= c
1521..2198
/*tag= c
EP-374869-A.
27-JUN-1990.
PD 1-JAN-1989; 123526.
PF 22-DEC-1978; GB-049807.
PR 27-DEC-1978; GB-050039.
PR 01-NOV-1979; GB-037910.
PA (BIOJ) Biogen Inc.
PI Murray K, Schaller HE;
DR WPI: 90-195067/26.
P-PSDB: R05634, R0535 & R06613.
DR Recombinant DNA encoding Hepatitis B virus polypeptide antigens -
PT used in detection of infection and in vaccine prodn.
PS Disclosure; English
CC Polypeptide fragments encoded by the DNA sequence are antigenic for
CC HBV and may be used as a vaccine or in detection. Peptides may be
CC cultured in a suitable bacterial host such as E.coli.
SQ Sequence 2743 BP; 638 A; 728 C; 579 G; 797 T; 1 Others;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 2743;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCCGAG 10
DB 279 GTCCTCCCGAG 270

RESULT 8
ID N81654
AC N81654:
DR 09-NOV-1990 (first entry)
DE LVDJ regions of human sperm-immobilising monoclonal antibody.
KW Anti-human sperm-immobilising monoclonal antibody; leader region;
KW vaccine; contraceptive; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds
622..912
/*tag= a
/product=variable region
FT misc_feature
528..610
/*tag= b
/*label=leader_region
FT misc_feature
622..912
/*tag= c
/*label=variable_region
```



```
FT misc_feature 913..943
FT /*tag= d
FT /label-diversity-region
FT misc_feature 944..995
FT /*tag= e
FT /label-joining-region
PN J63126482-A.
PD 30-MAY-1988.
PE 15-NOV-1986; 272412.
PR 15-NOV-1986; JP-272412.
PA (TOFU) TOA NENRIO KOGYO KK.
P-PSDB; P81259.
DR P-PSDB; P81259.
PT Cell strain producing human sperm-immobilising monoclonal antibody -
PT has at least v-gene in H chain coding genetic family and v gene in L
PT chain of DNA originating from antibody of sterile woman.
PS Disclosure; ; 14pp; Japanese.
CC A 1.85 kb IVDJ segment is inserted into pSV2gpt together with a
CC constant gamma 1 region. A cell strain, pref. a myeloma cell, can
CC produce the Ab with immobilising value (5150) of at least 5000,
CC sperm agglutination value at least 1:1600 dilution, specifically
CC reacting against human ejaculated sperm. The product can be used
CC as a vaccine and contraceptive.
SQ See also N81654-N81656.
SQ Sequence 1071 BP; 250 A; 268 C; 281 G; 272 T;

Query Match 100.0%; Score 10; DB 1; Length 1071;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCCTCCAG 10
DB 986 GTCCTCCAG 995

RESULT 9
006280/C
ID 006280 standard; DNA; 1399 BP.
AC 006280;
DE 03-FEB-1991 (first entry)
DE Sequence encoding human laminin B1 chain polypeptide fragment.
KW Metastasis; cancer; cell growth factor; tac promoter; ds.
OS Homo sapiens.
PN J02234679-A.
PD 17-SEP-1990.
PE 07-MAR-1989; 052835.
PE 07-MAR-1989; JP-052835.
(SAGA) SAGAMI CHEM RES CENTRE.
WPI; 90-325616/43.
P-PSDB; R07447.
DR Human laminin B1 chain polypeptide fragment, and expression
PT vector for inhibition of tumour metastasis, wound remedy
PT diagnostic agent etc
PS Claim 1; Fig 1; 20pp; Japanese.
CC Polypeptide may be produced from a transformed expression system
CC under the control of a tac promoter/operator, an SD sequenc of
CC metaprotease and tag ig gene.
CC The gene product has cell binding affinity, and may be used in
CC inhibition of metastasis and infection, wound healing, cell growth and
CC as a reagent for laminin assay.
SQ Sequence 1399 BP; 291 A; 376 C; 412 G; 320 T;

Query Match 100.0%; Score 10; DB 1; Length 1399;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCCTCCAG 10
DB 1155 GTCCTCCAG 1146
```

```
RESULT 10
011415/C
ID 011415 standard; cDNA; 15464 BP.
AC 011415;
DE 17-JUN-1991 (first entry)
DE Ryanodine receptor gene.
KW Malignant hyperthermia; hypermetabolic syndrome; inhalation;
KW anaesthetics; probe; calcium release channel; sarcoplasmic;
KW reticulum; ss.
OS Homo sapiens.
FH Key
FH Location/Qualifiers
FT cds
FT 105..15319
FT /*tag= a
FT /product- ryanodine receptor
FT 15439..15446
FT /*tag= b
FT 15447..15457
FT /*tag= c
FT /label- TG-rich region
FT 15464
FT /*tag= d
FT 328..773
FT /label- HRR-7B
FT /note= "0.4 kb; claim 6"
FT 400..2620
FT /*tag= f
FT /label- HRR-7A
FT /note= "2.2 kb; claim 6"
FT 2615..6359
FT /*tag= g
FT /label- HRR-5
FT /note= "3.8 kb; claim 6"
FT 6349..8717
FT /*tag= h
FT /label- HRR-4
FT /note= "2.4 kb; claim 6"
FT 9773..11075
FT /*tag= i
FT /label- HRR-3-1200
FT /note= "1.2 kb; claim 6"
FT 8738..9778
FT /*tag= j
FT /label- HRR-3-1000
FT /note= "1.0 kb; claim 6"
FT 8739..11842
FT /*tag= k
FT /label- HRR-3
FT /note= "3.1 kb; claim 6"
FT 11837..13831
FT /*tag= l
FT /label- HRR-2
FT /note= "2.0 kb; claim 6"
FT 13826..15467
FT /*tag= m
FT /label- HRR-1
FT /note= "1.6 kb; claim 6"
FT W09104328-A.
PN 04-APR-1991.
PD 21-SEP-1991; CA0312.
PE 25-SEP-1989; US-612726.
PR (HSCR-) HSC R & D Partnership.
PA (TORO) Univ. Toronto Innova.
PA (TORO-) Toronto Hospital.
PI Morton RG, MacLennan DH, Britt BA;
PI WPI; 91-117517/16.
P-PSDB; R11510.
DR Purified DNA specific for human ryanodine receptor - useful for
PT diagnosis of malignant melanoma.
PS Claim 1; Fig 2; 49pp; English.
CC The sequence was obtd. from several overlapping clones isolated
CC from a human skeletal muscle cDNA library in lambda gt10. The gene
CC (located on chromosome 19) encodes the human ryanodine receptor
```

CC (HRR), a calcium release channel which spans the gap between the  
 CC transverse tubule and the sarcoplasmic reticulum (SR) in muscle.  
 CC The sustained muscle contraction in malignant hyperthermia (MH)  
 CC may be caused by the release of calcium into the muscle cell cyto-  
 CC plasm from the SR. This is due to a defect in the gene encoding  
 CC HRR. Sequences from the gene can be used for diagnosis of MH  
 CC using RFLP analysis (see tags e-m).  
 SO Sequence 15464 BP; 3286 A; 4616 C; 4735 G; 2827 T;

Query Match 100.0%; Score 10; DB 1; Length 15464;  
 Best Local Similarity 100.0%; Pred. No. 9.6e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCCGAG 10  
 DB 2884 GTCCTCCCGAG 2875

.SULF 11

ID N91627 standard; DNA; 1196 BP.

AC N91627;

DT 19-MAR-1991 (first entry)

DE Sequence of gamma-3 intron switch sequences for immunoglobulin

DE genes.

DE Immunoglobulin gene switch sequence; intron switch sequence; ss.

OS Homo sapiens.

FT Key Location/Qualifiers

FT misc-feature 21..224

FT /tag- a

FT /label-sigma-gamma-3

FT misc-feature 225..1196

FT /tag- b

FT /label-sigma-gamma-4

PN MO8903425-A.

PD 20-APR-1989.

PF 14-OCT-1988; J01044.

PR 09-JUL-1988; JP-170041.

PA (EDUC-) Education Found Fujita (MITR MITR).

PI Kuromawa Y, Away A, Jshizuka Y, Wakabayash T;

DR WPT; 89-130043/17.

PT Recombinant DNA switch sequences for immunoglobulin genes - enable

PT specific control of prodn. of specific classes of antibody.

PS Disclosure: Fig 9B; 79pp; Japanese.

CC Sigma-gamma-3 and -4 are switch sequences for Cx genes upstream,

CC encoding human IgX (X - IgG3, IgG1, Ig-gamma-E, IgM, IgG2, IgG4,

CC IgE or IgA2). The DNA sequences are obtd. from suitable human cell

CC line, e.g. marrow cells, and pref. ARA-10 or Daudi, using a probe

CC for the IgH region, and cloned using a vector such as a plasmid of

CC (pref.) lambda-phase.

SO Sequence 1196 BP; 276 A; 312 C; 345 G; 263 T;

Query Match 100.0%; Score 10; DB 1; Length 1196;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCCGAG 10  
 DB 880 GTCCTCCCGAG 889

RESULT 12

Q11707/c

ID Q11707 standard; DNA; 14928 BP.

AC Q11707;

DT 25-JUL-1991 (first entry)

DE Immunoglobulin D-region heavy chain major cluster.

KW lymphoid tumour diagnosis; human; Ig; diversity region; ss.

OS Homo sapiens.

FT Key Location/Qualifiers

FT misc-rna 776..862

FT /tag- a

FT /label-Xp1 region

FT 1924..1995

FT /tag- b

FT /label-A4 region

FT 2885..2960

FT /tag- c

FT /label-K4 region

FT 4734..4807

FT /tag- d

FT /label-N4 region

FT 5238..5310

FT /tag- e

FT /label-M1 region

FT 7921..8007

FT /tag- f

FT /label-LR1 region

FT 10447..10533

FT /tag- g

FT /label-Xp1 region

FT 10631..10696

FT /tag- h

FT /label-Xp1 region

FT 11522..11593

FT /tag- i

FT /label-A1 region

FT 12478..12556

FT /tag- j

FT /label-K1 region

FT 13983..14059

FT /tag- k

FT /label-N1 region

FT 14490..14562

FT /tag- l

FT /label-M2 region

FT 1575..1829

FT /tag- m

FT /label-Region I

FT 3536..4394

FT /tag- n

FT /label-Region II

FT 12686..13184

FT /tag- o

FT /label-Region III

FT 13232..13247

FT /tag- p

FT /label-Region IV

PN J03087185-A.

PD 11-APR-1991.

PF 19-DEC-1989; 329005.

PR 19-DEC-1988; JP-319809.

PR 19-JUN-1989; JP-154623.

PR 19-DEC-1989; JP-329005.

PA (MITR ) MITSUI TOATSU CHEM INC.

DR WPT; 91-152426/21.

PT Human immunoglobulin gene - for diagnosis and monitoring of

PT lymphoid tumour esp. in leukaemia

PS Claim 2; Fig 2; 35pp; Japanese.

CC The sequence is a 15kb fragment of the D(H) region of the human Ig

CC gene (between D(LR4) and D(LR2)). See also Q11708.

SO Sequence 14928 BP; 3308 A; 4995 C; 4066 G; 2559 T;

Query Match 100.0%; Score 10; DB 1; Length 14928;  
 Best Local Similarity 100.0%; Pred. No. 9.6e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCCGAG 10  
 DB 6743 GTCCTCCCGAG 6734

RESULT 13



```

FT      /label= ATTRA motif
FT      /note= "reduces mRNA stability"
FT      4240.4244
FT      /tag= q
FT      /label= ATTRA motif
FT      /note= "reduces mRNA stability"
FT      4249.4253
FT      /tag= r
FT      /label= ATTRA motif
FT      /note= "reduces mRNA stability"
FT      4280.4284
FT      /tag= s
FT      /label= ATTRA motif
FT      /note= "reduces mRNA stability"
FT      4357.4362
FT      /tag= t
FT      poly_a_signal
FT      4357.4362
FT      WO9110738-A.
FT      25-JUL-1991.
FT      08-JAN-1991; US-462158.
FT      08-JAN-1990; US-462158.
FT      (RESE ) RES CORP TECHN INC.
FT      Vansnick J, Uytendhoe C, Simpson RJ;
FT      WPI; 91-238026/32.
FT      P-P-SDB; R13217.
FT      Nucleic acid encoding for P40 T-cell growth factor - for patients
FT      with AIDS or compromised immune systems, also allows increased
FT      prodn. of other cytokine(s).
FT      PS      Claim 2; Fig 14; 122pp; English.
FT      CC      The sequence was obtd. from a clone, lambda H40.3a1, isolated from
FT      CC      a genomic library using a murine P40 CDNA clone as a probe. The
FT      CC      sequence, or the cDNA sequence obtd. using the sequence, can be
FT      CC      inserted into a vector for expression of P40 in a host organism.
FT      CC      The protein is useful for stimulating the proliferation of certain
FT      CC      subsets of T helper cells e.g. in AIDS patients or immune
FT      CC      compromised patients.
FT      CC      See also Q12760, and Q13660-Q13663.
FT      SO      Sequence 4382 BP, 1210 A; 922 C; 904 G; 1346 T;

```

```

Query Match      100.0%; Score 10; DB 1; Length 4382;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 GTCCTCCCGAG 10
      |||||
      367 GTCCTCCCGAG 376

```

```

Search completed: September 18, 1999, 00:33:39
Job time: 18979, sec

```

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:06 : Search time 2825.05 Seconds  
(without alignments)  
6.982 Million cell updates/sec

Title: US-08-956-518a-93  
Perfect score: 10  
Sequence: 1 GTCCTCCAG 10

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	10	100.0	200	20	D20867	D20867 HUMS01848
C 2	10	100.0	161	20	D25774	D25774 HUMS04142
C 3	10	100.0	166	20	D28621	D28621 MDS0601 MO
C 4	10	100.0	424	20	D32336	D32336 CELK01154R
C 5	10	100.0	360	20	D35681	D35681 CELK02487F
C 6	10	100.0	360	20	D36081	D36081 CELK02883F
C 7	10	100.0	418	20	D37701	D37701 CELK00808F
C 8	10	100.0	391	20	D41072	D41072 RICS3337A R
C 9	10	100.0	324	20	F02878	F02878 HSC1ED092 n
C 10	10	100.0	268	20	F03210	F03210 HSC1ED082 n
C 11	10	100.0	521	20	L26678	L26678 MDSF080A 1a
C 12	10	100.0	401	20	L26715	L26715 MDS167A 1am
C 13	10	100.0	309	20	L34981	L34981 HUMZ02EST H
C 14	10	100.0	397	20	M78691	M78691 EST00839 H1
C 15	10	100.0	338	20	M85776	M85776 EST02297 Fe
C 16	10	100.0	411	20	M89223	M89223 CEL1868 Chr
C 17	10	100.0	340	20	T00993	T00993 WEST01714 E
C 18	10	100.0	236	20	T00994	T00994 WEST01715 E
C 19	10	100.0	352	20	T03019	T03019 FB20D6 Feta
C 20	10	100.0	213	20	T03520	T03520 IB403 Infan
C 21	10	100.0	199	20	T03860	T03860 seq2338 4HF
C 22	10	100.0	238	20	T04845	T04845 EST02733 Fe
C 23	10	100.0	289	20	T06308	T06308 EST04197 Fe
C 24	10	100.0	374	20	T06613	T06613 EST04502 Fe
C 25	10	100.0	343	20	T07835	T07835 EST05725 Fe
C 26	10	100.0	379	20	T10329	T10329 seq1249 b4H
C 27	10	100.0	1069	20	T12278	T12278 EST0003 Sub
C 28	10	100.0	146	20	T12345	T12345 A044F Heart
C 29	10	100.0	397	20	T15603	T15603 IB1616 Infia
C 30	10	100.0	250	20	T16832	T16832 NIB1910-5R
C 31	10	100.0	450	20	T18269	T18269 5C01905-t7
C 32	10	100.0	408	20	T19397	T19397 f02001s Tes
C 33	10	100.0	397	20	T24183	T24183 CTS171 1amB
C 34	10	100.0	363	20	T27108	T27108 NIBT290C05R
C 35	10	100.0	340	20	T27693	T27693 EST12419 Hu
C 36	10	100.0	213	20	T28192	T28192 EST131471 Hu
C 37	10	100.0	253	20	T29572	T29572 EST185052 Hu
C 38	10	100.0	271	20	T30400	T30400 EST15995 Hu
C 39	10	100.0	269	20	T30791	T30791 EST12689 Hu
C 40	10	100.0	448	20	T33051	T33051 EST156484 Hu
C 41	10	100.0	364	20	T33660	T33660 EST158625 Hu
C 42	10	100.0	345	20	T35723	T35723 EST190125 Hu
C 43	10	100.0	345	20	T36837	T36837 EST101849 S
C 44	10	100.0	307	20	T37940	T37940 EST103255 S
C 45	10	100.0	490	54	HSN009637	A1044787 Homo sapi

## ALIGNMENTS

RESULT 1  
D20867/c D20867 200 bp mRNA EST  
LOCUS HUMS01848 Human promyelocyte Homo sapiens cDNA clone mp224 3',  
DEFINITION mRNA sequence.  
D20867  
ACCESSION G504687  
NID D20867.1 GI:504687  
VERSION

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
TITLE 1 (bases 1 to 200)  
JOURNAL Murakawa, K., Matsubara, K., Fukushima, A., Yoshii, J. and Okubo, K.  
MEDLINE Chromosomal assignments of 3'-directed partial cDNA sequences  
95137584 representing novel genes expressed in granulocytoid cells  
Genomics 23, 379-389 (1994)

2 (bases 1 to 200)  
Okubo, K., Fukushima, A., Yoshii, J., Niiyama, T., Kojima, Y.,  
Yoshinari, H., Arimoto, J. and Matsubara, K.  
gene expression of human promyelocytic cell line HL60 before and  
after induction of differentiation. A new application of 3'-directed  
cDNA sequencing  
Unpublished (1993)

JOURNAL  
COMMENT

CONTACT: Okubo, K., Fukushima, A., Yoshii, J., Niiyama, T., Kojima, Y.,  
Yoshinari, H., Arimoto, J. and Matsubara, K.  
Institute for Molecular and Cellular Biology  
Osaka University  
3-1 Yamada-oka, Suita, Osaka 565, Japan.  
Location/Qualifiers

FEATURES  
SOURCE  
1. .200  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="6"  
/clone\_lib="mp2234"  
/note="Female, adult, cell\_line = HL60, cell\_type =  
promyelocyte."  
promyelocyte.

BASE COUNT 55 a 58 c 40 g 42 t 5 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 200;  
Best Local Similarity 100.0%; Pred. No. 8.4e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCAG 10  
|||||  
Db 27 GTCCTCCAG 18

RESULT 2  
5774/c  
-JUS 161 bp mRNA EST 30-NOV-1995  
DEFINITION HUMS04142 Human colon mucosa Homo sapiens cDNA clone cm0231 3',  
mRNA sequence.  
ACCESSION D25774  
NID 9500461  
VERSION D25774.1 GI:500461  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
TITLE 1 (bases 1 to 161)  
Okubo, K., Yoshii, J., Yokouchi, H., Kamayama, M. and Matsubara, K.  
global analysis of gene expression in colon mucosa: a large scale  
random cDNA sequencing analysis  
Unpublished (1994)

JOURNAL  
COMMENT

CONTACT: Okubo, K., Itoh, K., Yoshii, J., Yokouchi, H. and Matsubara, K.  
Institute for Molecular and Cellular Biology  
Osaka University  
3-1 Yamada-oka, Suita, Osaka 565, Japan.  
Location/Qualifiers

FEATURES  
SOURCE  
1. .161  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

/clone="cm0231"  
/clone\_lib="Human colon mucosa"  
/note="Adult male, tissue\_type = colon mucosa"

BASE COUNT 47 a 31 c 37 g 46 t

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 161;  
Best Local Similarity 100.0%; Pred. No. 7.9e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCAG 10  
|||||  
Db 26 GTCCTCCAG 17

RESULT 3  
D28621/c  
LOCUS D28621 166 bp mRNA EST 07-OCT-1996  
DEFINITION M0586G01 mouse embryonal carcinoma cell line F9 Mus musculus cDNA  
clone 86G01, mRNA sequence.  
ACCESSION D28621  
NID 9518938  
VERSION D28621.1 GI:618938  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
TITLE 1 (bases 1 to 166)  
Nishiguchi, S., Joh, T., Horie, K., Zou, Z., Yasunaga, T. and Shimada, K.  
A survey of genes expressed in undifferentiated mouse embryonal  
carcinoma F9 cells: characterization of low-abundance mRNAs  
J. Biochem. 116, 128-139 (1994)  
95096008

JOURNAL  
MEDLINE  
COMMENT

CONTACT: Kazunori Shimada  
Department of Medical Genetics, Division of Molecular Biomedicine  
Research Institute for Microbial Diseases, Osaka University  
3-1, Yamadaoka, Suita, Osaka, 565, Japan  
Tel: 06-879-8325  
Fax: 06-879-8326.  
Location/Qualifiers

FEATURES  
SOURCE  
1. .166  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="mouse embryonal carcinoma cell line F9"  
/clone="86G01"

BASE COUNT 33 a 44 c 43 g 44 t 2 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 166;  
Best Local Similarity 100.0%; Pred. No. 8e+03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCAG 10  
|||||  
Db 20 GTCCTCCAG 11

RESULT 4  
D32336  
LOCUS D32336 424 bp mRNA EST 05-AUG-1994  
DEFINITION CELK011BAR Yujii Kohara unpublished cDNA Caenorhabditis elegans cDNA  
clone YK11B4 3', mRNA sequence.  
ACCESSION D32336  
NID 9522953  
VERSION D32336.1 GI:522953  
KEYWORDS EST.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdittia; Rhabdittida;

REFERENCE 1 (bases 1 to 424)  
 AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and Tabara,H.  
 TITLE Toward an expression map of the C.elegans genome  
 JOURNAL Unpublished (1994)  
 COMMENT

Contact: Yuji Kohara  
 Gene Library Lab  
 National Institute of Genetics  
 Yata 1111, Mishima, Shizuoka 411, Japan  
 Tel: 0559-75-0771  
 Fax: 0559-75-6240  
 Email: ykohara@dbj.nig.ac.j.  
 Location/Qualifiers

FEATURES  
 source 1..424  
 /organism="Caenorhabditis elegans"  
 /strain="CB1489 him-8(e1489)"  
 /db\_xref="taxon:6239"  
 /clone="YK11b4"  
 /clone\_1lb="Yuji Kohara unpublished cDNA"  
 /note="dev\_stage=varied, sex=Hermaphrodite male, tissue\_type=whole animal"

BASE COUNT 150 a 78 c 83 g 112 t 1 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 424;  
 Best Local Similarity 100.0%; Pred. No. 1e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCCGAG 10  
 |||||||||  
 Db 337 GTCCTCCCGAG 346

RESULT 5  
 D35681/c 360 bp mRNA EST 08-AUG-1994  
 LOCUS CELK02487F Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA  
 DEFINITION clone yk24b7 5', mRNA sequence.  
 ACCESSION D35681  
 NID 9527131  
 VERSION D35681.1 GI:527131  
 KEYWORDS EST.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 Rhabditina; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 360)  
 AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and Tabara,H.  
 TITLE Toward an expression map of the C.elegans genome  
 JOURNAL Unpublished (1994)  
 COMMENT

Contact: Yuji Kohara  
 Gene Library Lab  
 National Institute of Genetics  
 Yata 1111, Mishima, Shizuoka 411, Japan  
 Tel: 0559-75-0771  
 Fax: 0559-75-6240  
 Email: ykohara@dbj.nig.ac.j.  
 Location/Qualifiers

FEATURES  
 source 1..360  
 /organism="Caenorhabditis elegans"  
 /strain="CB1489 him-8(e1489)"  
 /db\_xref="taxon:6239"  
 /clone="YK24b7"  
 /clone\_1lb="Yuji Kohara unpublished cDNA"  
 /note="dev\_stage=varied, sex=Hermaphrodite male, tissue\_type=whole animal"

BASE COUNT 144 a 49 c 88 g 78 t 1 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 360;  
 Best Local Similarity 100.0%; Pred. No. 9.7e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCCGAG 10  
 |||||||||  
 Db 202 GTCCTCCCGAG 193

RESULT 6  
 D36081/c 360 bp mRNA EST 08-AUG-1994  
 LOCUS CELK028H3F Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA  
 DEFINITION clone yk28h3 5', mRNA sequence.  
 ACCESSION D36081  
 NID 9528019  
 VERSION D36081.1 GI:528019  
 KEYWORDS EST.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 Rhabditina; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 360)  
 AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and Tabara,H.  
 TITLE Toward an expression map of the C.elegans genome  
 JOURNAL Unpublished (1994)  
 COMMENT

Contact: Yuji Kohara  
 Gene Library Lab  
 National Institute of Genetics  
 Yata 1111, Mishima, Shizuoka 411, Japan  
 Tel: 0559-75-0771  
 Fax: 0559-75-6240  
 Email: ykohara@dbj.nig.ac.j.  
 Location/Qualifiers

FEATURES  
 source 1..360  
 /organism="Caenorhabditis elegans"  
 /strain="CB1489 him-8(e1489)"  
 /db\_xref="taxon:6239"  
 /clone="YK28h3"  
 /clone\_1lb="Yuji Kohara unpublished cDNA"  
 /note="dev\_stage=varied, sex=Hermaphrodite male, tissue\_type=whole animal"  
 BASE COUNT 96 a 86 c 124 g 51 t 3 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 360;  
 Best Local Similarity 100.0%; Pred. No. 9.7e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCCGAG 10  
 |||||||||  
 Db 248 GTCCTCCCGAG 239

RESULT 7  
 D37701/c 418 bp mRNA EST 08-AUG-1994  
 LOCUS CELK008D8F Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA  
 DEFINITION clone yk8d8 5', mRNA sequence.  
 ACCESSION D37701  
 NID 9526082  
 VERSION D37701.1 GI:526082  
 KEYWORDS EST.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 Rhabditina; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 418)

BASE COUNT 144 a 49 c 88 g 78 t 1 others  
 ORIGIN

AUTHORS Kohara, Y., Mitsuki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and Tabara, H.  
 TITLE Toward an expression map of the C.elegans genome  
 JOURNAL Unpublished (1994)  
 COMMENT

FEATURES  
 source  
 Contact: Yuji Kohara  
 Gene Library Lab  
 National Institute of Genetics  
 Yata 1111, Mishima, Shizuoka 411, Japan  
 Tel: 0559-75-0771  
 Fax: 0559-75-6240  
 Email: ykohara@dbj.nig.ac.jp.  
 Location/Qualifiers  
 1.418  
 /organism="Caenorhabditis elegans"  
 /strain="CB1489 him-8(e1489)"  
 /db\_xref="taxon:6239"  
 /clone="Yk868"  
 /clone\_lib="Yuji Kohara unpublished cDNA"  
 /note="dev\_stage=varied, sex=Hermaphrodite male, tissue\_type=whole animal"  
 /tissue\_type="whole animal"

BASE COUNT 104 a 94 c 99 g 121 t  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 418;  
 Best Local Similarity 100.0%; Pred. No. 1e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTCCTCCCG 10  
 |||||||||  
 Db 395 GTCCTCCCG 386

RESULT 8  
 D41072 391 bp mRNA EST 15-NOV-1994  
 LOCUS R1C33337A Rice shoot Oryza sativa cDNA, mRNA sequence.  
 DEFINITION D41072  
 ACCESSION D41072  
 NID 9571561  
 VERSION D41072.1 GI:571561  
 KEYWORDS EST.  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
 Poaceae; Oryza.  
 1 (bases 1 to 391)  
 Sasaki, T., Miyao, A. and Yamamoto, K.  
 Rice cDNA from callus 1995  
 Unpublished (1995)  
 COMMENT  
 Contact: Takuji Sasaki  
 National Institute of Agricultural Resources  
 Rice Genome Research Program  
 2-1-2 Kamondai, Tsukuba  
 Ibaraki,  
 Japan 305  
 Tel: 0298-38-7441  
 Fax: 0298-38-7468  
 Email: tsasaki@dbj.nig.ac.jp.  
 Location/Qualifiers  
 1.391  
 /organism="Oryza sativa"  
 /strain="Nipponbare sub-species Japonica"  
 /db\_xref="taxon:4530"  
 /clone\_lib="Rice shoot"  
 /note="Etiolated shoot (8 days old)"

BASE COUNT 72 a 140 c 119 g 54 t 6 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 391;

Best Local Similarity 100.0%; Pred. No. 9.9e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTCCTCCCG 10  
 |||||||||  
 Db 67 GTCCTCCCG 76

RESULT 9  
 F02878/c 324 bp mRNA EST 02-FEB-1995  
 LOCUS HSC1E092 normalized infant brain cDNA Homo sapiens cDNA clone  
 DEFINITION c-1ed09 3', mRNA sequence.  
 ACCESSION F02878  
 NID 9646435  
 VERSION F02878.1 GI:646435  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 324)  
 Auffray, C., Behar, G., Bols, F., Bouchier, C., da Silva, C.,  
 Devignes, M.D., Duprat, S., Houligatte, R., Jumeau, M.N., Lamy, B.,  
 Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pletu, G., Pouliot, Y.,  
 Sebastiani, K., Kitchin, C. and Tessier, A.  
 IMAGE: molecular integration of the analysis of the human genome  
 and its expression  
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
 9527534  
 JOURNAL  
 MEDLINE  
 COMMENT

CONTACT: Genethon  
 Genexpress-Genethon  
 Genethon Centre de recherche sur le Genome Humain  
 1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE  
 Tel: 33169472800  
 Fax: 3316078698  
 Email: genexpress@genethon.fr  
 Single read. removed at sequence 5' end  
 Genexpress\_library\_id: C; Genexpress\_sequence\_id: alc-1ed09  
 Seq primer: (-21)M13 universal.  
 Location/Qualifiers  
 1.324  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="c-1ed09"  
 /clone\_lib="normalized infant brain cDNA"  
 /sex="Female"  
 /tissue\_type="total brain"  
 /dev\_stage="3 months old"  
 /note="Organ: brain; Vector: lambda BD; Site: 1, HindIII;  
 Site 2: NotI; sex: Female; dev\_stage=3 months old;  
 isolate=muscular atrophy patient; tissue\_type=total  
 brain; total mRNA was oligo-(dT) primed and directionally  
 cloned 5' -> 3' into the HindIII -> NotI sites of the  
 lambda BD vector. Clone library from B. Soares, psychiatry  
 Dept. Columbia University, USA. Normalization method:  
 Bento Soares, P.N.A.S. in press"

BASE COUNT 67 a 84 c 119 g 53 t 1 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 324;  
 Best Local Similarity 100.0%; Pred. No. 9.4e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTCCTCCCG 10  
 |||||||||  
 Db 228 GTCCTCCCG 219

RESULT 10  
 F03210/c



LOCUS F03210 268 bp mRNA EST 02-FEB-1995  
 DEFINITION HSC1G082 normalized infant brain cDNA Homo sapiens cDNA clone  
 c-1P908 3', mRNA sequence.  
 ACCESSION F03210  
 NID 9646767  
 VERSION F03210.1 GI:646767  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 268)  
 AUTHORS Aulifay,C., Benar,G., Bois,F., Bouchier,C., da Silva,C.,  
 Devignes,M.D., Duprat,S., Houligatte,R., Jumeau,M.N., Lamy,B.,  
 Lorenzo,F., Mitchell,H., Marigalte-Sanson,R., Pletu,G., Pouillot,Y.,  
 Sebastiani,Kabackich,C. and Tessier,A.  
 IMAGE: molecular integration of the analysis of the human genome  
 and its expression  
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
 95277534  
 JOURNAL MEDLINE  
 MEDLINE  
 CONTACT: Genethon  
 Genexpress-Genethon  
 Genethon Centre de recherche sur le Genome Humain  
 1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE  
 Tel: 33169472800  
 Fax: 33160778698  
 Email: genexpress@genethon.fr  
 Single read. removed at sequence 5'end  
 Genexpress\_library\_id: C; Genexpress\_sequence\_id: alc-1p908  
 Seq primer: (-21)M13.universal.  
 FEATURES  
 SOURCE  
 1. 268  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="c-1p908"  
 /clone\_1db="normalized infant brain cDNA"  
 /sex="Female"  
 /tissue\_type="total brain"  
 /dev\_stage="3 months old"  
 /note="Organ: brain; Vector: lambda BA; Site\_1: HindIII;  
 Site\_2: NotI; sex:Female; dev\_stage=3 months old;  
 isolate=muscular atrophy patient; tissue\_type=total  
 brain; total mRNA was oligo-(dT) primed and directionally  
 cloned 5' -> 3' into the HindIII -> NotI sites of the  
 lambda BA vector. Clone library from B.Souares, Psychiatry  
 Dept. Columbia University, USA. Normalization method:  
 Bento Soares, P.N.A.S in press"  
 BASE COUNT 74 a 53 c 71 g 67 t 3 others  
 ORIGIN  
 Query Match 100.0%; Score 10; DB 20; Length 268;  
 Best Local Similarity 100.0%; Pred. NO. 9e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTCCTCCCGAG 10  
 |||||||||  
 Db 119 GTCTCCCGAG 110  
 RESULT 11  
 LOCUS 126678 521 bp mRNA EST 01-MAR-1995  
 DEFINITION MUS0F080A lambda unizap male testis Mus musculus cDNA clone F080.  
 mRNA sequence.  
 ACCESSION 126678  
 NID 9437469  
 VERSION 126678.1 GI:437469  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE 1 (bases 1 to 521)  
 AUTHORS Kerr,S.M., Vambrie,S., McKay,S.J. and Cooke,H.J.  
 TITLE Analysis of cDNA sequences from mouse testis  
 JOURNAL Mamm. Genome 5, 557-565 (1994)  
 MEDLINE 95093181  
 COMMENT Contact: Kerr,S.M., Vambrie,S., McKay,S.J. and Cooke,H.J.  
 FEATURES  
 SOURCE  
 1. 521  
 /organism="Mus musculus"  
 /strain="Swiss"  
 /db\_xref="taxon:10090"  
 /clone="F080"  
 /clone\_1db="lambda unizap male testis"  
 /note="Vector: lambda unizap; lambda unizap library from  
 mall mouse testis "  
 BASE COUNT 131 a 127 c 129 g 132 t 2 others  
 ORIGIN  
 Query Match 100.0%; Score 10; DB 20; Length 521;  
 Best Local Similarity 100.0%; Pred. NO. 1.1e+04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTCCTCCCGAG 10  
 |||||||||  
 Db 216 GTCTCCCGAG 225  
 RESULT 12  
 LOCUS 126715 401 bp mRNA EST 01-MAR-1995  
 DEFINITION MUS167A lambda unizap male testis Mus musculus cDNA clone F167,  
 mRNA sequence.  
 ACCESSION 126715  
 NID 9437393  
 VERSION 126715.1 GI:437393  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 401)  
 AUTHORS Kerr,S.M., Vambrie,S., McKay,S.J. and Cooke,H.J.  
 TITLE Analysis of cDNA sequences from mouse testis  
 JOURNAL Mamm. Genome 5, 557-565 (1994)  
 MEDLINE 95093181  
 COMMENT Contact: Kerr,S.M., Vambrie,S., McKay,S.J. and Cooke,H.J.  
 FEATURES  
 SOURCE  
 1. 401  
 /organism="Mus musculus"  
 /strain="Swiss"  
 /db\_xref="taxon:10090"  
 /clone="F167"  
 /clone\_1db="lambda unizap male testis"  
 /note="Vector: lambda unizap; lambda unizap library from  
 mall mouse testis "  
 BASE COUNT 113 a 87 c 124 g 71 t 6 others  
 ORIGIN  
 Query Match 100.0%; Score 10; DB 20; Length 401;  
 Best Local Similarity 100.0%; Pred. NO. 9.9e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTCCTCCCGAG 10  
 |||||||||  
 Db 274 GTCTCCCGAG 265  
 RESULT 13

L34981  
 LOCUS L34981 309 bp mRNA EST 29-DEC-1994  
 DEFINITION HUM20EST Human brain striatum Homo sapiens cDNA clone Y20, mRNA sequence.  
 ACCESSION L34981  
 NID 9522224  
 VERSION L34981.1 GI:522224  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 309)  
 AUTHORS Jou, Y.-S., Gould, R.D. and Myers, R.M.  
 TITLE Localization of the alpha 2-macroglobulin receptor-associated protein 1 gene (LRPAP1) and other gene fragments to human chromosome 4p16.3 by direct cDNA selection  
 JOURNAL Genomics 24, 410-413 (1994)  
 MEDLINE 95213044  
 COMMENT  
 FEATURES  
 source Contact: Jou, Y.-S., Gould, R.D. and Myers, R.M.  
 Location/Qualifiers  
 1.309  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="4p16.3"  
 /clone="Y20"  
 /clone\_lib="Human brain striatum"  
 /note="tissue\_type = brain striatum"  
 BASE COUNT 38 a 84 c 87 g 94 t 6 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 309;  
 Best Local Similarity 100.0%; Pred. No. 9.5e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCCTCCCG 10  
 Db 1 GTCCTCCCG 10

RESULT 14  
 M78691 397 bp mRNA EST 26-MAY-1992  
 LOCUS M78691/c  
 DEFINITION EST00839 Hippocampus, Striatum (cat. #936205) Homo sapiens cDNA clone HRCMC15, mRNA sequence.  
 ACCESSION M78691  
 NID 9273006  
 VERSION M78691.1 GI:273006  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 397)  
 AUTHORS Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C.  
 TITLE Sequence identification of 2,375 human brain genes  
 JOURNAL Nature 355, 632-634 (1992)  
 MEDLINE 92168112  
 COMMENT  
 FEATURES  
 source Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org  
 Seq primer: M13 Reverse.  
 Location/Qualifiers  
 1.337  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

/db\_xref="ATCC (inhost):78098"  
 /db\_xref="GDB:DS1679E"  
 /db\_xref="taxon:9606"  
 /clone="HRCMC15"  
 /clone\_lib="Hippocampus, Striatum (cat. #936205)"  
 /note="Vector: LambdaZAP-II; Female, 2 years; oligo-dT + random primed cDNA synthesis; LambdaZAP-II vector, 1.0kb average insert size."  
 BASE COUNT 104 a 106 c 96 g 88 t 3 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 397;  
 Best Local Similarity 100.0%; Pred. No. 9.9e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCCTCCCG 10  
 Db 289 GTCCTCCCG 280

RESULT 15  
 M85776 338 bp mRNA EST 11-JAN-1995  
 LOCUS M85776/c  
 DEFINITION EST02297 Fetal brain, Striatum (cat#936206) Homo sapiens cDNA clone HRCMC15, mRNA sequence.  
 ACCESSION M85776  
 NID 9274425  
 VERSION M85776.1 GI:274425  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 338)  
 AUTHORS Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C.  
 TITLE Sequence identification of 2,375 human brain genes  
 JOURNAL Nature 355, 632-634 (1992)  
 MEDLINE 92168112  
 COMMENT  
 REFERENCE 2 (bases 1 to 338)  
 AUTHORS Durdin, A.S.  
 TITLE EST mapping data  
 JOURNAL Unpublished (1996)  
 COMMENT  
 FEATURES  
 source Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org  
 Seq primer: M13 Forward.  
 Location/Qualifiers  
 1.338  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):81353"  
 /db\_xref="GDB:DS581E"  
 /db\_xref="taxon:9606"  
 /map="8"  
 /clone="HRCMC15"  
 /clone\_lib="Fetal brain, Striatum (cat#936206)"  
 /note="Vector: LambdaZAP-II; 17-18 wk gestation, female; oligo-dT + random primed cDNA synthesis; LambdaZAP-II vector, 1.0kb average insert size."  
 BASE COUNT 85 a 65 c 94 g 92 t 2 others  
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 338;  
 Best Local Similarity 100.0%; Pred. No. 9.5e+03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

sep 19 10:27:03 1999

us-08-956-518a-93.rst

Page 7

OY 1 GTCCTCCAG 10  
|||||  
Db 44 GTCCTCCAG 35

Search completed: September 17, 1999, 21:28:07  
Job time: 14271 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:01:46 ; Search time 1962.17 Seconds

(without alignments)  
740.711 Million cell updates/sec

Title: US-08-956-518a-94

Sequence: 1 AGAAGCAGGAGGAGGAGTAG.....CTCTGACAGGTAAGCCAC 457

Scoring table: IDENTITY\_NUC

Searched: 679419 segs, 1590154680 residues

Database :

GenEmbl: \*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_to:\*  
13: gb\_st:\*  
14: gb\_sts:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_vl:\*  
18: em\_fun:\*  
19: em\_hlg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_om:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_to:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_vl:\*  
34: gb\_hlg1:\*  
35: gb\_hlg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_ba1:\*  
39: em\_ba2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_p14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	392	85.8	392	11	AF029837 Homo sapi

2	127	27.8	1876	10	HSU62436	U62436 Human nicot
3	121.6	26.6	2087	10	HSAR7A	X70297 H. sapiens m
4	73.6	16.1	168919	35	AC007686	AC007686 Homo sapi
5	70	15.3	342379	34	CEY39B6	Z55399 Caenorhabdi
6	68	14.9	1072	17	HS21E	M29384 Herpes simp
7	65.8	14.4	1560	10	HS21E	M29384 Herpes simp
8	63.2	13.8	43058	10	HS21E	M29384 Herpes simp
9	63.2	13.8	2685	10	HUMBR3	J00172 Human DNA S
10	63	13.8	154746	17	HSV2HG52	Z86099 Herpes simp
11	63	13.8	154746	17	HSV2HG52	Z86099 Herpes simp
12	62	13.6	1977	10	HSU40583	U40583 Human alpha
13	61.6	13.5	38532	10	SC994	AD009628 Streptomy
14	61.6	13.5	41936	10	CH19R30879	AD009628 Streptomy
15	61.6	13.5	14440	10	HSUSP2	Y07661 H. sapiens U
16	61.4	13.4	3252	10	HSUSP2	Y11354 H. sapiens m
17	61.4	13.4	210636	11	HSATFI13	AC006443 Homo sapi
18	61.2	13.4	248287	35	AC007225	AC007225 Homo sapi
19	61	13.3	1927	3	CEGTPBPAM	Z12168 C. famillari
20	61	13.3	557	4	XELRGE12	K01371 X. laevis oo
21	61	13.3	3924	4	XELRGE12	J00999 X. laevis ex
22	61	13.3	7634	4	XELRGE12	X59734 X. laevis 28
23	61	13.3	8133	4	XLRN01	X02995 Xenopus lae
24	59	12.9	634	14	G40486	G40486 26971 Zebra
25	58.2	12.7	163974	9	AB011399	AB011399 Homo sapi
26	58.2	12.7	163974	9	AB011399	AB011399 Homo sapi
27	58.2	12.7	331211	10	AB016897	AB016897 Homo sapi
28	58.2	12.7	147971	10	HS431P23	AL009178 Human DNA
29	58.2	12.7	179436	35	AC003066	AC003066 Mus muscu
30	57.8	12.6	181884	34	HS35C18	AL022327 Homo sapi
31	57.6	12.6	240229	35	AC003059	AC003059 Mus muscu
32	57.2	12.5	1027	12	MMIGFITE2	X71919 M. musculus
33	57.2	12.5	267925	35	AC007228	K01371 X. laevis oo
34	57	12.5	557	4	XELRGE12	J00999 X. laevis ex
35	57	12.5	3924	4	XELRGE12	X59734 X. laevis 28
36	57	12.5	7634	4	XLRN01	X02995 Xenopus lae
37	57	12.5	8153	4	XLRN01	AC007224 Homo sapi
38	56.8	12.4	204263	35	AC007224	AC007224 Homo sapi
39	56.4	12.3	39213	11	AC003047	AC003047 Homo sapi
40	56.4	12.3	152261	17	HEICG	X14112 Herpes simp
41	56.4	12.3	152261	17	HEICG	X14112 Herpes simp
42	56.4	12.3	6633	17	HERSVIG3	X06461 Herpes simp
43	56.4	12.3	26245	17	HSIUS	L00036 HSV1 (stra)
44	56.4	12.3	26245	17	HSIUS	L00036 HSV1 (stra)
45	56.2	12.3	1766	17	HS1LS17	K01835 HSV-1 (stra

## ALIGNMENTS

## RESULT 1

AF029837

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

AF029837 392 bp DNA PRI 16-DEC-1998  
Homo sapiens alpha 7 neuronal nicotinic receptor gene, promoter  
region.  
AF029837  
G3757792  
AF029837.1 GI:3757792  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 392)  
Gault,J., Robinson,M., Berger,R., Drebing,C., Logel,J., Hopkins,J.,  
Moore,T., Jacobs,S., Meriwether,J., Chol,M.J., Kim,E.J., Walton,K.,  
Bulfinch,K., Davis,A., Breese,C., Freedman,R. and Leonard,S.  
Genomic organization and partial duplication of the human alpha7  
neuronal nicotinic acetylcholine receptor gene  
Genomics 52 (2), 173-185 (1998)  
9900837  
2 (bases 1 to 392)  
Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R.,  
Breese,C., Davis,A., Hopkins,J. and Freedman,R.

TITLE		Direct Submission	
JOURNAL		Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA	
FEATURES		Location/Qualifiers	
SOURCE		1. 392	
PROMOTER		/organism="Homo sapiens"	
BASE COUNT		/db_xref="taxon:9606"	
ORIGIN		/chromosome="15"	
		/map="15q14; between D1S1031 and D1S1360"	
		/clone="CHRNA7-P"	
		1. 392	
		51 a 122 c 180 g 39 t	
Query Match		85.8%; Score 392; DB 11; Length 392;	
Best Local Similarity		100.0%; Pred. No. 2.7e-46;	
Matches 392; Conservative		0; Mismatches 0; Indels 0; Gaps	
QY	1	AGAACCGAAGGAGAGGTAGAGCTGGCTTTGGGACACCCCTGGCTGGCCAGAGGCGG	60
DB	1	AGAACGCAAGAGGAGGTAGAGCTGGCTTTGGGACACCCCTGGCTGGCCAGAGGCGG	60
QY	61	AGGCGGAGAGCCGCTGGGTGAGACACCTGGGGGCTGGAGTCCCGGAGCCGACCAAGCC	120
DB	61	AGGCGGAGAGCCGCTGGGTGAGACACCTGGGGGCTGGAGTCCCGGAGCCGACCAAGCC	120
QY	121	GGGAGTACCTCCCGCTCAACACCTGGGGCTGAGTTCCTGGGTGGCCGCGAGACGCTGG	180
DB	121	GGGAGTACCTCCCGCTCAACACCTGGGGCTGAGTTCCTGGGTGGCCGCGAGACGCTGG	180
QY	181	CCCGGGCTGAGAGGATGGCGGGGCGGGGACGAGGCGGGGCGGGGCTCTCAAGTGGAG	240
DB	181	CCCGGGCTGAGAGGATGGCGGGGCGGGGACGAGGCGGGGCGGGGCTCTCAAGTGGAG	240
QY	241	AGCGCGCGGGGGGCGGGGCGGGGCGGGGCGGGGCTCTTAAGGCGGCGGAGCC	300
DB	241	AGCGCGCGGGGGGCGGGGCGGGGCGGGGCGGGGCTCTTAAGGCGGCGGAGCC	300
QY	301	GAGGGGAGAGGTGCTCTGTGGCGGAGGGGCGAGGCCCGGGCGACAGCGAGAGTGGAG	360
DB	301	GAGGGGAGAGGTGCTCTGTGGCGGAGGGGCGAGGCCCGGGCGACAGCGAGAGTGGAG	360
QY	361	CGCGCCGGCTCGCTGACAGCTCCGGGACTCAAC	392
DB	361	CGCGCCGGCTCGCTGACAGCTCCGGGACTCAAC	392
RESULT	2		
LOCUS	HSU62436	1876 bp	11-JAN-1997
DEFINITION	Human nicotinic acetylcholine receptor alpha7 subunit precursor,		
ACCESSION	U62436		
NID	91458119		
VERSION	062436.1		
KEYWORDS	GI:1458119		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	1 (bases 1 to 1876)		
TITLE	Elliot,R.J., Ellis,S.B., Berchman,K.J., Urrutia,A.,		
	Chavez-Noriega,L.E., Johnson,E.C., Velicelebi,G. and Harpold,M.M.,		
	Comparative structure of human neuronal alpha 2-alpha 7 and beta		
	2-beta 4 nicotinic acetylcholine receptor subunits and functional		
	expression of the alpha 2, alpha 3, alpha 4, alpha 7, beta 2, and		
	beta 4 subunits		
	J. Mol. Neurosci. 7 (3), 217-228 (1996)		
MEDLINE	97062879		
REFERENCE	2 (bases 1 to 1876)		
AUTHORS	Elliot,R.J.		
TITLE	Direct Submission		

JOURNAL Submitted (28-JUN-1996) Kathryn J. Elliott, SIBIA Neurosciences, Inc., 505 Coast Blvd. So., La Jolla, CA 92037, USA

FEATURES  
SOURCE Location/Qualifiers  
1. 1876  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Kalpha7.3"  
/clone\_id="SIBIA.Lambda gtl1 library #3 (M. Williams)"  
/cell\_line="IMR32"  
/cell\_type="neuroblastoma"  
1. 172  
73. 138  
73. 1581

5' UTR  
sig\_peptide  
CDS  
/codon\_start=1  
/product="nicotinic acetylcholine receptor alpha7 subunit precursor"  
/protein\_id="AAB40114.1"  
/db\_xref="PID:91458120"  
/db\_xref="GI:1458120"  
/translation="MRCSPGCVWALALASLIHVSLOGEFQKLYKELKYNPLERP  
AMDSQPLTVYFSLIQIMVDERNOVLTNIMLOMSMDTHYLIQMNSEYGVTVR  
PDQIMKPDILLVNSADERPATEHTVNVNSGHCYLLPGIRKSSCYIDVRWPEPF  
VQHKLFKFSWYSGVGLDIOMPADISGYIPNEMDPLVGIPIGRSEFYCKEPEY  
DMETVMMRRRTLYVYGINLIPCYLISALALVPLPADSGEKISGITVTLSTVE  
LVAAEIPATSDSVPPLAQYFASMTITVGLSVYTVYLYQTHHDDPDGKAPKMTRV  
LLNCAWFLLRKRPGEDEVKRPACQKHQRCSLASVEVAAPPAISNGLYIGRGI  
DGWVCVPTPDGCVVCGMACSPTEDEHLHSGQPEGSDPLAKLLEEVRYIANFRRC  
DESAVCSWENKFAVCVDRICLMAFSVFTICTIGILMSAPNFEAVASKDPA"  
139. 1578  
/product="nicotinic acetylcholine receptor alpha7 subunit  
1582. 1876

BASE COUNT 369 a 553 c 531 g 423 t  
ORIGIN

Query Match 27.8%; Score 127; DB 10; Length 1876;  
Best Local Similarity 100.0%; Pred. No. 5.9e-10;  
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 321 GGGCGAGGCGGAGGCGGAGCGAGAGCGTGAGCGCGCGGCGTGGCTCAGCT 380  
Db 1 GGGCGAGGCGGAGGCGGAGCGAGAGCGTGAGCGCGCGGCGTGGCTCAGCT 60  
Y 381 CCGGACTCAACATCGCTGCTGCGCGGAGGCGCTGCTGCGCTGGCCGCGTGCCTC 440  
Db 61 CCGGACTCAACATCGCTGCTGCGCGGAGGCGCTGCTGCGCTGGCCGCGTGCCTC 120  
Y 441 CTGCACG 447  
111111  
Db 121 CTGCACG 127

RESULT 3  
LOCUS HSNARA7A 2087 bp mRNA PRI 01-JUN-1994  
DEFINITION H.sapiens mRNA for neuronal nicotinic acetylcholine receptor  
alpha-7 subunit.  
ACCESSION X70297  
NID 9486606  
KEYWORDS X70297.1. GI:496606  
SOURCE neuronal nicotinic acetylcholine receptor alpha-7 subunit.  
human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 2087)  
PENG, X., Katz, M., Gerzanich, V., Anand, R., and Lindstrom, J.  
Human alpha 7 acetylcholine receptor: cloning of the alpha 7  
subunit from the SH-SY5Y cell line and determination of  
pharmacological properties of native receptors and functional alpha  
7 homomers expressed in Xenopus oocytes  
Mol. Pharmacol. 45 (3), 546-554 (1994)

JOURNAL MEDLINE  
94195283

REFERENCE 2 (bases 1 to 2087)  
 AUTHORS Katz, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-FEB-1993) M. Katz, Univ of Pennsylvania School of Medicine, Dept of Neuroscience, 36th & Hamilton Walk, Rm 235 Stemmler Hall, Philadelphia, PA 19104, USA  
 FEATURES  
 SOURCE  
 1. 2087  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /tissue\_type="brain"  
 /cell\_type="neuroblastoma"  
 /cell\_line="SHSY-5Y"  
 /clone\_id="lambda zap II"  
 /clone="SHSY3"  
 104..175  
 /product="neuronal nicotinic acetylcholine receptor  
 alpha-7 subunit"  
 104..1612  
 /product="neuronal nicotinic acetylcholine receptor  
 alpha-7 subunit"  
 /codon\_start=1  
 /product="neuronal nicotinic acetylcholine receptor  
 alpha-7 subunit"  
 /protein\_id="CAA49778.1"  
 /db\_xref="PIR:G496607"  
 /db\_xref="GI:496607"  
 /translation="MRCSPGVVGLAASLHYSLOGERORRLYKELVKNYPLNPV  
 ANDSPLIVFSLSLQIMVDKNOVLTNIMLQMSWIDHYLQMWSEYGVKTVRF  
 PQGLWPKPILLYNSADERDAFHNVLVNSSHCQYLPPIGKRSCTIDVMPFED  
 VQCKLKFGSWSYGSLDQMOEADISGIIPIGMEWDVIGIPGRSESECEKEYP  
 DVFETVMRRRTLYGLNLLPCVLISALALVELLPADSGEKISIGITLSTLTEM  
 LLYAEIMPATSDSPLIAQYFASTMTIVIGSVYTVIVLQYHNDPDGCGMPKTVYI  
 LNMCAWFLRMKRGEDKVRPAQHQKRSGLASVEMSAVGPAPPAANGNLTIGFEGGL  
 DEHYCVPTDSDGVCGRMASPTDHLHLHGQPPGSDPOLARILEVRIANRFGCO  
 DSEAVCSSEMKRAACVYDRICLMAFSEVFTIICITIGIMSNAPNEVAVSDFA"  
 mat\_peptide 441 a 601 c 573 g 472 t  
 BASE COUNT 176..1609  
 ORIGIN

Query Match 26.6%; Score 121.6; DB 10; Length 2087;  
 Best Local Similarity 96.1%; Pred. No. 3.2e-09;  
 Matches 146; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 296 GAGCCGAGCGCCGAGAGCTGTGTGCGCGAGCGCGCGGAGACCGGAGCG 355  
 |||||||  
 Db 9 GAGCCGAGCGCGAGAGCTGTGTGCGCGAGCGCGCGGAGACCGGAGCG 66  
 |||||||

QY 356 TGGAGCGCGCGGCTGCTGCTGCTGCGGAGCTGCTGCTGCTGCTGCTGCT 415  
 |||||||  
 Db 67 TGGAGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 126  
 |||||||

QY 416 CTGGCTGGCGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 447  
 |||||||  
 Db 127 CTGGCTGGCGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 158  
 |||||||

RESULT 4  
 AC007686 168919 bp DNA HTG 30-MAY-1999  
 LOCUS Homo sapiens chromosome 14 clone BAC 2289B16 map 14q24.3, LOW-PASS  
 DEFINITION SEQUENCE SAMPLING.  
 AC007686  
 NID 94927297  
 VERSION AC007686.1 GI:4927297  
 KEYWORDS HTG; HTGS\_PHASE0.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 168919)  
 James, R., Loretz, C., Lasky, S., Madan, A., Prescott, S., Ratcliffe, A.,

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 COMMENT  
 TITLE  
 JOURNAL  
 COMMENT

Shaffer, T. and Hood, L.  
 Sequencing of human chromosome 14q24.3 region  
 2 (bases 1 to 168919)  
 Rowen, L., Madan, A., Qin, S., Abbasi, N., Dors, M., Dickhoff, R.,  
 James, R., Loretz, C., Lasky, S., Madan, A., Prescott, S., Ratcliffe, A.,  
 Shaffer, T. and Hood, L.  
 Direct Submission  
 Submitted (30-MAY-1999) Multimegabase Sequencing Center, University  
 of Washington, PO Box 357730, Seattle, WA 98195, USA  
 \* NOTE: his record contains 192 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1 965: contig of 965 bp in length  
 gap of unknown length  
 966 1866: contig of 901 bp in length  
 gap of unknown length  
 1867 2747: contig of 881 bp in length  
 gap of unknown length  
 2748 3608: contig of 861 bp in length  
 gap of unknown length  
 3609 4462: contig of 854 bp in length  
 gap of unknown length  
 4463 5345: contig of 883 bp in length  
 gap of unknown length  
 5346 6187: contig of 842 bp in length  
 gap of unknown length  
 6188 7072: contig of 885 bp in length  
 gap of unknown length  
 7073 7927: contig of 855 bp in length  
 gap of unknown length  
 7928 8797: contig of 870 bp in length  
 gap of unknown length  
 8798 9647: contig of 850 bp in length  
 gap of unknown length  
 9648 10555: contig of 908 bp in length  
 gap of unknown length  
 10556 11387: contig of 832 bp in length  
 gap of unknown length  
 11388 12285: contig of 898 bp in length  
 gap of unknown length  
 12286 13144: contig of 859 bp in length  
 gap of unknown length  
 13145 14010: contig of 866 bp in length  
 gap of unknown length  
 14011 14869: contig of 855 bp in length  
 gap of unknown length  
 14870 15756: contig of 887 bp in length  
 gap of unknown length  
 15757 16611: contig of 855 bp in length  
 gap of unknown length  
 16612 17482: contig of 871 bp in length  
 gap of unknown length  
 17483 18330: contig of 848 bp in length  
 gap of unknown length  
 18331 19206: contig of 876 bp in length  
 gap of unknown length  
 19207 20072: contig of 866 bp in length  
 gap of unknown length  
 20073 20946: contig of 874 bp in length  
 gap of unknown length  
 20947 21791: contig of 845 bp in length  
 gap of unknown length  
 21792 23944: contig of 2153 bp in length  
 gap of unknown length

23945 24793: contig of 849 bp in length  
24794 25663: contig of 870 bp in length  
25664 26504: contig of 841 bp in length  
26505 27381: contig of 877 bp in length  
27382 28254: contig of 873 bp in length  
28255 29126: contig of 872 bp in length  
29127 29964: contig of 859 bp in length  
30822 30822: contig of 858 bp in length  
30823 31681: contig of 858 bp in length  
31681 32538: contig of 858 bp in length  
32539 33390: contig of 852 bp in length  
33391 34255: contig of 865 bp in length  
34256 35105: contig of 850 bp in length  
35106 36030: contig of 925 bp in length  
36031 36869: contig of 839 bp in length  
36870 37734: contig of 865 bp in length  
37735 38584: contig of 850 bp in length  
38585 39448: contig of 864 bp in length  
39449 40283: contig of 835 bp in length  
40284 41169: contig of 886 bp in length  
41170 42014: contig of 845 bp in length  
42015 42927: contig of 913 bp in length  
42928 43788: contig of 861 bp in length  
43789 44650: contig of 862 bp in length  
44651 45491: contig of 841 bp in length  
45492 46383: contig of 892 bp in length  
46384 47232: contig of 849 bp in length  
47233 48096: contig of 864 bp in length  
48097 48951: contig of 855 bp in length  
48952 49831: contig of 880 bp in length  
49832 50676: contig of 845 bp in length  
50677 51534: contig of 858 bp in length  
51535 52388: contig of 854 bp in length  
52389 53266: contig of 878 bp in length  
53267 54108: contig of 842 bp in length  
54109 55136: contig of 1028 bp in length  
55137 55986: contig of 850 bp in length

55987 56874: gap of unknown length  
56875 57721: gap of unknown length  
57722 58587: gap of unknown length  
58588 59427: gap of unknown length  
59428 60292: gap of unknown length  
60293 61149: gap of unknown length  
61150 62022: gap of unknown length  
62023 62858: gap of unknown length  
62859 63733: gap of unknown length  
63734 64579: gap of unknown length  
64580 65451: gap of unknown length  
65452 66289: gap of unknown length  
66290 67174: gap of unknown length  
67175 68020: gap of unknown length  
68021 68903: gap of unknown length  
68904 69757: gap of unknown length  
69758 70628: gap of unknown length  
70629 71467: gap of unknown length  
71467 72361: gap of unknown length  
72362 73206: gap of unknown length

Query Match 16.1%; Score 73.6; DB 35; Length 168919;  
Best Local Similarity 49.5%; Pred. No. 0.0031;  
Matches 190; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

53 GAGGCGGAGGCGGAGACCCGCTGCTGAGAGCTGGGGGGGAGAGTCCCGGAGACCTAC 112  
42534 GGGGGGGGGGGCGGG 42533  
113 CCAGGCGGGGGAGTACCTCCCGCTCACACCTCGGGCTGACATTCCTGGTGGCCGCCGA 172  
42594 CGGGGGCGGTGGCGGG 42593  
173 GACGCTGCGCCGCTGAGAGGATGCGGGGGGGGAGCGGGGGCGGGGGCGGGGCTGTC 232  
42654 GGGGGGTGTCGG 42713  
233 ACGTGAGAGCGCGCGCGGG 292  
42714 GGGGGGGGGGGCGGG 42773  
293 CGCGAGCGGAGGGGAGGAGTGTCTGTGGCCGAGCGGAGCGGCGGCGGAGCGGAG 352  
42774 GCGGGGGCGGG 42833  
353 ACGTGAGCGCGCGCGCTGCTGACAGCTCGGGAGCTCAACATGCGTCTCCCGGGAGG 412  
42834 GCGGGGGGGGAGGG 42893  
413 CGTGTGCTGGCGCTGGCGCGCTG 436  
42894 GGGCGGGCGGG 42917

RESULT 5  
CEY39B6/c  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

## MENT

Direct Submission  
Submitted (09-JUN-1999) Nematode Sequencing Project, Sanger Centre,  
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,  
Washington University, St. Louis, MO 63110, USA. E-mail:  
jess@sanger.ac.uk or twenematode.wustl.edu  
On Jun 11, 1999 this sequence version replaced gi:4725958.  
Order of segments is not known; 800 n's separate segments.  
IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence. Work on the sequence is in progress  
and the release of this data is based on the understanding that the  
sequence may change as work continues. The sequence may be  
contaminated with foreign sequence from E.coli, yeast, vector,  
phage etc.  
\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

FEATURES  
source  
Location/Qualifiers  
1.342379  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/chromosome="IV"  
/clone="Y3986"

BASE COUNT 106570 a 59268 c 58805 g 109715 t 8021 others  
ORIGIN

Query Match 15.3%; Score 70; DB 34; Length 342379;  
Best Local Similarity 50.9%; Pred. No. 0.0076;  
Matches 166; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 43 GGCTGTGCGAGAGCGCGAGAGCGCGCTGCTGAGTGGAGTGGAGTGGC 102  
Db 330422 GGG 330363  
QY 103 CGAGAGCTACCAAGCGCGGAGTACCTCCGCTACACCTCGGCTGCAATTCCCTGGG 162  
Db 330362 GGG 330303  
QY 163 TGGCGCGGAGAGCGTGGCGCGCGGCTGAGAGTGGCGGCGGAGCGGCGGCG 222  
Db 330302 AGG 330243  
QY 223 GGGGCTGCTACGTGAGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGCTC 282  
Db 330242 GGG 330183  
QY 283 CTTAAAGCGCGCGAGCGGAGCGAGTCTCTGTGCGCGCAGCGCGAGCGCGCGC 342  
Db 330182 GG 330123  
QY 343 GACAGCGGAGAGCGTGGAGCGCGCGCGCG 368  
Db 330122 GCCGGGGGGGGGGGGGGGGAGCGGGGG 330097

RESULT 6

HS2IEA  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
FEATURES

Location/Qualifiers  
1.1072  
/organism="human herpesvirus 2"  
/db\_xref="taxon:10310"  
prim\_transcript 95..>1072  
/note="IE-5 mRNA and Intron"  
344..866  
/note="IE-5 Intron"  
897..>1072  
/note="Immediate early protein 5"  
/codon\_start=1  
/protein\_id="AAA45849.1"  
/db\_xref="PID:G330284"  
/db\_xref="GI:330284"  
/translation="MSMALKTDMPSDSRCTHRYGVDCAEIHKREREDPAATAV  
TDEPLPLDPPVPRP"

BASE COUNT 132 a 322 c 543 g 75 t  
ORIGIN

Query Match 14.9%; Score 68; DB 17; Length 1072;  
Best Local Similarity 50.6%; Pred. No. 0.098;  
Matches 164; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 37 GCCCTGCGCTGCGCAAGAGCGCGAGCGCGCGCTGCTGAGTGGAGTGGG 96  
Db 366 GCCGGGGGAGACGGCGCGGGGGGAGCGGGCGGAGCGGGCGGCGGGGG 425  
QY 97 GTGCGCGGAGCTACCAAGCGCGGAGTACCTCCGCTACACCTCGGCTGCAATT 156  
Db 426 GACGGGCGCGGGGAGCGCGCGGGGGGAGCGGGCGGAGCGGGCGGCGGGG 485  
QY 157 CTTGGGTGCGCGCGAGAGCTGCTGAGAGTGGCGGGCGGAGCGGGGGC 216  
Db 486 GCCGGGGGCGGGGGGAGCGGGGGGAGCGGGCGGCGGGCGGCGGGGGGA 545  
QY 217 GGGGCGGGGCTGCTACGTGAGAGCGCGCGGGGGCGGGGGCGGGGGCGCGCGC 276  
Db 546 CGGGGGGAGCGGGGGGAGCGGGGGGAGCGGGGGCGGGGGGAGCGGGGGCGG 605  
QY 277 CGGCTCTTAAAGCGCGGAGCGCGGAGCGAGTCTCTGTGCGCGCAGCGCGAGC 336  
Db 606 GGGGAGCGGGCGGGGGGCGGGGGGAGCGGGCGGGGGGAGCGGGGGCGGGGGG 665  
QY 337 CGGGGGAGAGCGGAGAGCTGGAG 360  
Db 666 CGGGGGGAGCGGGGGGAGCGGGGGG 689

RESULT 7

HS2IE  
LOCUS  
DEFINITION

HS2IE 1560 bp DNA VRL 15-MAR-1990  
Herpes simplex virus type 2 Immediate-early (IE4) protein mRNA, 5'



end.  
 M29384  
 NID 9330281  
 VERSION M29384.1 GI:330281  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Herpes simplex virus type 2 (strain HGS Timbury, 1971) DNA, passed in baby hamster kidney clone 21 (Cl3) cells.  
 human herpesvirus 2  
 Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 Alphaherpesvirinae; Simplexvirus.  
 1 (bases 1 to 1560)  
 Whitton, J.L. and Clements, J.B.  
 The junctions between the repetitive and the short unique sequences of the herpes simplex virus genome are determined by the polypeptide-coding regions of two spliced immediate-early mRNAs  
 J. Gen. Virol. 65, 451-466 (1984)  
 84137573  
 MEDLINE  
 JOURNAL  
 FEATURES  
 source  
 Location/Qualifiers  
 1..1560  
 /organism="human herpesvirus 2"  
 /db\_xref="taxon:10310"  
 prim\_transcript 95..>1560  
 /note="IE-4 mRNA and intron"  
 Intron 344..906  
 /note="IE-4 intron"  
 CDS 969..>1560  
 /note="immediate early protein 4"  
 /codon\_start=1  
 /protein\_id="AA04584.1"  
 /db\_xref="GI:555152"  
 /db\_xref="PIR:G55152"  
 /translation="MADIPDPDPALDTTPANHAPSPPLPGSRKRRPYLPSSSESGK  
 PDIESSTESSEDEVDLKGRRRPRELGLDSESTGTESEGTGSDDD  
 DDDASDGLVDPKSKRPRLNLTSPDRGVVPEVWRSDRPIRAAQPAPQ  
 SSGDAHAPRRAQRQMRSAATLDTLHYRQCNQ"  
 BASE COUNT 222 a 480 c 714 g 144 t  
 ORIGIN  
 Query Match 14.4% Score 65.8; DB 17; Length 1560;  
 Best Local Similarity 51.3%; Pred. No. 0.17;  
 Matches 177; Conservative 0; Mismatches 167; Indels 1; Gaps 1;  
 25 TGGCCCTTGGGCAACCCCTGCGCTGAGAGGCGCGAGACCCCGCTCGTGAGAG 84  
 350 TGAACCGGCGCGCGCGCGGAGAGCGCGCGGAGAGCGCGCGCGCGCGCGG 409  
 85 ACTGGGGGTGGAGGTCCCGGAGCGTACCCAGCGCGGAGATGACTCCGCTCACACTC 144  
 410 ACGGCGCGGAGGAGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 469  
 145 GGGCTCAGTCCCTGGTGGCGCGCGGAGCGCTGCGCGCGGTGAGGATGGCGCGCG 204  
 470 GGGGCGCGGCG 529  
 205 GGGGAGCGGCGCGCGCGCGCGCGCTCTACGTGAGAGCGCGCGCGCGCGCGCG 264  
 529 GGGGCGCGGCG 588  
 265 GGGGCGCGCGCGCGCGCTCTTAAGCGCGCGCGAGCGCGCGAGTGCCTCTGTGGCC 324  
 589 GCGCGCGCGCGCGCGCGGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 648  
 325 GCAAGCGCGCGCGCGCGCGCGCGAGCGTGTGAGCGCGCGCGCGCGCGCG 369  
 649 GGGAGCGGCG 693  
 RESULT 8  
 LOCUS HSGG1 43058 bp DNA PRI 19-MAR-1997  
 DEFINITION Human DNA sequence from cosmid GGI from a contig from the tip of the short arm of chromosome 16, spanning 2Mb of 16p13.3 Contains

alpha and zeta globin genes and ESTs.  
 284721  
 NID 91817575  
 VERSION 284721.1 GI:1817575  
 KEYWORDS 16p13.3; alpha-globin; globin; zeta-globin.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 43058)  
 Flint, J. and Higgs, D.R.  
 Direct Submission  
 Submitted (28-JAN-1997) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1RQ, UK. E-mail enquiries: humpdesanger.ac.uk  
 IMPORTANT: This sequence is the entire insert of clone GGI. This  
 clone was sequenced at the Institute of Molecular Medicine. The  
 true left end of clone GGI is at 1 in this sequence. The true right  
 end of clone GGI is at 43058.  
 GGI is from a 280kb clone contig extending from the telomere of  
 16p. Higgs D.R., Flint J. unpublished. MRC Molecular Haematology  
 Unit, Institute of Molecular Medicine, Oxford.  
 GGI came from the Los Alamos, flow sorted human Chromosome 16  
 library.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..43058  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /map="16p13.3"  
 /clone="GGI"  
 1..72  
 /note="FRAM repeat: matches 98..166 of consensus"  
 97..399  
 /note="Alusg repeat: matches 1..300 of consensus"  
 516..811  
 /note="AlusC repeat: matches 1..294 of consensus"  
 823..977  
 /note="L1 repeat: matches 5163..5328 of consensus"  
 904..1028  
 /note="L1ME3A repeat: matches 2..146 of consensus"  
 1109..1760  
 /note="L1MB8 repeat: matches 134..862 of consensus"  
 1763..2057  
 /note="AluY repeat: matches 2..297 of consensus"  
 2443..2607  
 /note="AluY repeat: matches 297..133 of consensus;  
 incomplete repeat"  
 2608..2905  
 /note="AluY repeat: matches 297..1 of consensus"  
 4231..4328  
 /note="MIR repeat: matches 182..85 of consensus"  
 4691..4989  
 /note="L1 repeat: matches 7..301 of consensus"  
 5222..5528  
 /note="Alusg repeat: matches 1..300 of consensus"  
 5530..5698  
 /note="Alusg repeat: matches 130..296 of consensus;  
 incomplete repeat"  
 5714..6184  
 /note="L1MB5 repeat: matches 922..459 of consensus"  
 6320..6503  
 /note="L1MB7 repeat: matches 2..208 of consensus"  
 6531..6825  
 /note="Alusg repeat: matches 295..1 of consensus"  
 6864..7002  
 /note="MSTC repeat: matches 1..136 of consensus"  
 7045..7351  
 /note="AluY repeat: matches 1..301 of consensus"  
 7447..7545  
 /note="MSTB repeat: matches 294..394 of consensus"  
 7545..7885  
 /note="L1MB3 repeat: matches 229..569 of consensus"  
 7979..8281

```

Repeat_region
/note="AluX repeat: matches 1. .302 of consensus"
8296. .8597
Repeat_region
/note="AluY repeat: matches 1. .301 of consensus"
8635. .8924
Repeat_region
/note="AluJb repeat: matches 12. .299 of consensus"
8928. .8997
Repeat_region
/note="L1ME3 repeat: matches 631. .700 of consensus"
9052. .9327
Repeat_region
/note="AluY repeat: matches 13. .294 of consensus"
10382. .10428
Repeat_region
/note="AluX/g repeat: matches 245. .291 of consensus;
incomplete repeat"
10652. .10713
Repeat_region
/note="L1MB8 repeat: matches 920. .859 of consensus"
10714. .10966
Repeat_region
/note="AluSp repeat: matches 1. .255 of consensus;
incomplete repeat"
11022. .11168
Repeat_region
/note="AluSg repeat: matches 1. .145 of consensus;
incomplete repeat"
11163. .11463
Repeat_region
/note="AluSp repeat: matches 1. .296 of consensus"
11466. .11635
Repeat_region
/note="AluXg repeat: matches 133. .302 of consensus;
incomplete repeat"
11730. .12317
Repeat_region
/note="L1 repeat: matches 4729. .5317 of consensus"
12321. .12616
Repeat_region
/note="AluY repeat: matches 1. .296 of consensus"
12633. .12761
Repeat_region
/note="AluSg repeat: matches 1. .131 of consensus;
incomplete repeat"
12779. .13073
Repeat_region
/note="AluSg repeat: matches 1. .297 of consensus"
13011(13773. .13867,14755. .14955,15135. .15263)
/codon_start=1
/product="zeta-globin 1"
/protein_id="CAB06552.1"
/db_xref="PID:6300370"
/db_xref="PID:61817576"
/db_xref="GI:1817576"
/db_xref="SWISS-PROT:P02008"
/translation="WSLTKEITITIVSMANISIQADTIGTETLEFLSHPQRTYF
PFHDPASQALRAHGSKVAVAGVADVISIDIGALSLSLHAYILRDVPMFKLL
SHCLVLTAAAREPAFTAEAAHAKMKFLSVSVSLTEYKR"
13923. .14090
Repeat_region
/note="12 copies of 14 mer 89 & conserved"
14618. .15730
Repeat_region
/note="Putative CpG island"
14964. .15071
Repeat_region
/note="3 copies of 36 mer 83 & conserved"
15747. .15875
Repeat_region
/note="MIR2 repeat: matches 145. .10 of consensus"
16241. .16541
Repeat_region
/note="AluY repeat: matches 301. .1 of consensus"
16892. .17027
Repeat_region
/note="AluUo repeat: matches 1. .136 of consensus;
incomplete repeat"
17029. .17326
Repeat_region
/note="AluY repeat: matches 1. .297 of consensus"
17368. .17506
Repeat_region
/note="AluUo repeat: matches 10. .151 of consensus;
incomplete repeat"
17945. .18086
Repeat_region
/note="AluUo repeat: matches 153. .294 of consensus;
incomplete repeat"
18106. .18426
Repeat_region
/note="AluUo repeat: matches 1. .301 of consensus"
18715. .18833
Repeat_region
/note="L1MB7 repeat: matches 744. .865 of consensus"
18834. .19130
Repeat_region
/note="AluXg repeat: matches 1. .296 of consensus"
19186. .19449

```

	repeat_region	/note="ME42c repeat: matches 1506. .1220 of consensus" 19431. .19593 /note="HWM11 repeat: matches 1221. .1058 of consensus" 19601. .19838 /note="Alusg repeat: matches 300. .63 of consensus; incomplete repeat" 19840. .19983 /note="Alusg repeat: matches 154. .300 of consensus; incomplete repeat" 20709. .20856 /note="4 copies of 37 mer 93 % conserved" 20881. .20952 /note="2 copies of 36 mer 99 % conserved" 20919. .21028 /note="3 copies of 37 mer 97 % conserved" 20996. .21067 /note="2 copies of 36 mer 99 % conserved" 21039. .21186 /note="4 copies of 37 mer 95 % conserved" 21216. .21323 /note="3 copies of 36 mer 97 % conserved" 21325. .21396 /note="2 copies of 36 mer 97 % conserved" 21360. .21433 /note="2 copies of 37 mer 97 % conserved" 21398. .21469 /note="2 copies of 36 mer 97 % conserved" 21471. .21542 /note="2 copies of 36 mer 99 % conserved" 21546. .21617 /note="2 copies of 36 mer 96 % conserved" 21579. .21726 /note="4 copies of 37 mer 94 % conserved" 22591. .22661 /note="Alusx repeat: matches 302. .232 of consensus; incomplete repeat" 22766. .23070 /note="Alusx repeat: matches 1. .302 of consensus" join(23985. .24079,25345. .25549,25891. .26019) /codon_start=1 /pseudo /product="zeta-globin 2" /db_xref="PDB:e300371" 24135. .24680 /note="39 copies of 14 mer 90 % conserved" 25208. .27584 /note="Putative CpG island" prim_transcript <26871..>27290 /note="match: multiple ESTs; similar to HEMOGLOBIN ALPHA-D; match: 5' EST H57003 clone .204625; match: 5' EST R92558 clone 196372; match: 3' EST H67154 clone 211156"
Query Match		13.8%; Score 63.2; DB 10; Length 43058;
Best Local Similarity	52.7%;	Fred. No. 0.13;
Matches 137; Conservative	0; Mismatches 123; Indels	0; Gaps 0
OY	24 CTTGGCTTGGAGACCCCTGGCTGCAGAGCGGAGCGGAGGCCCGCTGGGTGA	83
Db	25589 CGGGCGGGGCGGGGCGCGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGG	25648
OY	84 GACTGGGGGTGGAGTGCCCGGAGCGTAACCACCGCCCGGAGTACTCCGCTCACACT	143
Db	25649 GGGCGGGGTGGG	25708
OY	144 CGGGTCGACTTCCCTGGGTGGCCCCCGAGACGTGGCCCCGGGCTGGAGAGATGGCGGG	203
Db	25709 GGG	25768
OY	204 CGGGACGG	263
Db	25769 CCGGGGCGGGGTCCGG	25828
OY	264 GGGGGCGGGCGGGCGGGCTCC	283

Db 25829 GGGGTCTGGGGGGGGGGCC 25848

## RESULT 9

HUMBA3 2685 bp DNA PRI 18-SEP-1997  
 LOCUS Human alpha globin gene cluster on chromosome 16: psizeta  
 DEFINITION pseudogene.  
 ACCESSION J00184  
 J0183792  
 J00184.1 GI:183792  
 VERSION gene duplication; pseudogene; repeat region; zeta-globin.  
 KEYWORDS 3 of 4  
 SEGMENT  
 SOURCE Homo sapiens DNA.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 2685)  
 AUTHORS Proudfoot, N.J., Gill, A., and Maniatis, T.  
 TITLE The structure of the human zeta-globin gene and a closely linked,  
 nearly identical pseudogene  
 JOURNAL Cell 31 (3 Pt 2), 553-563 (1982)  
 MEDLINE 83129370  
 COMMENT The human alpha globin gene cluster, located on the short arm of  
 chromosome 16, spans about 30 kb and includes the following five  
 loci:  
 5'- zeta - pseudozeta - pseudalpha-1 - alpha-2 - alpha-1 -3'. The  
 gene sequence shown below is considered a pseudogene because of the  
 terminating codon at base 429. There are only two additional  
 single-base differences between the coding sequences of the zeta  
 and pseudozeta genes, although there are conspicuous differences in  
 the 5' and 3' flanking regions. Moreover the first intron of this  
 locus demonstrates length variation, apparently stemming from the  
 repeat sequence 'acagtgaggaggg' which is also found in the zeta  
 gene and, in a similar form, in the 5' flank of the human insulin  
 gene. The repeat sequence 'cgagg' is characteristic of the second  
 introns of the zeta genes [1].  
 Typical promoter elements 'caccat' and 'tata' are found at bases  
 291 and 326.

## FEATURES

source 1. .2685  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="16p13.3"  
 356. .505  
 /gene="phbz"  
 /number=1  
 join(411. .505,1767. .1971,2313. .2441)  
 /gene="phbz"  
 pseudo  
 /codon\_start=1  
 join(411. .505,1767. .1971,2313. .2441)  
 /gene="phbz"  
 429  
 /gene="phbz"  
 /note="c in phbz; g in hbx (see <humba3>)"  
 506. .1766  
 /gene="phbz"  
 /number=1  
 /note="repeat sequence copy 1"  
 1767. .1971  
 /gene="phbz"  
 /number=2  
 1972. .2312  
 /gene="phbz"  
 /number=2  
 1977. .1981  
 /note="repeat sequence, copy 1"  
 2313. .2541  
 /gene="phbz"  
 /number=3  
 exon  
 repeat\_region  
 intron  
 mutation  
 gene  
 cds  
 exon  
 source

polyA\_signal 2521. .2526  
 /note="phbz mRNA polyadenylation signal"  
 BASE COUNT 454 a 720 c 1155 g 356 t  
 ORIGIN About 1 kb after segment 2.

Query Match 13.8%; Score 63.2; DB 10; Length 2685;  
 Best Local Similarity 52.7%; Pred. No. 0.33;  
 Matches 137; Conservative 0; Mismatches 123; Indels 0; Gaps 0;  
 QY 24 CTGGCTTGGGACACCCCTGCGCTGCGCAGAGCGGCGGAGCGGCTGCTGGA 83  
 Db 2011 CGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 2070  
 QY 84 GACTGGGGGTGAGAGTCCCGGAGCTACCCAGCGCGGGAGTACCTCCGCTCACACT 143  
 Db 2071 GGGCGGGGTGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 2130  
 QY 144 CGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 203  
 Db 2131 GGGCGG 2190  
 QY 204 CGGGGACGGG 263  
 Db 2191 CGGGGCGGGGGGTGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 2250  
 QY 264 GGGGGCGGGCGGGCGGGCTCC 283  
 Db 2251 GGGGTCTCGGGGGGGGGGGCC 2270

## RESULT 10

HSVHG52 154746 bp DNA VRL 04-DEC-1998  
 LOCUS Herpes simplex virus type 2 (strain HG52), complete genome.  
 DEFINITION  
 ACCESSION 286099  
 NID 91869820  
 VERSION 286099.1 GI:1869820  
 KEYWORDS capsid associated tegument protein; capsid protein; complete genome; deoxyribonuclease; deoxyuridine triphosphatase; DNA polymerase; envelope protein; glycoprotein B; glycoprotein C; glycoprotein D; glycoprotein E; glycoprotein F; glycoprotein G; glycoprotein H; glycoprotein I; glycoprotein J; glycoprotein K; glycoprotein L; glycoprotein M; host shut-off factor; integral membrane protein; major capsid protein; membrane-associated phosphoprotein; minor capsid protein; neurovirulence factor; protease; protein kinase; ribonucleotide reductase large subunit; ribonucleotide reductase small subunit; RL1 gene; RL2 gene; RL3 gene; tegument protein; thymidine kinase; UL1 gene; UL10 gene; UL11 gene; UL12 gene; UL13 gene; UL14 gene; UL15 gene; UL16 gene; UL17 gene; UL18 gene; UL19 gene; UL2 gene; UL20 gene; UL21 gene; UL22 gene; UL23 gene; UL24 gene; UL25 gene; UL26 gene; UL26.5 gene; UL27 gene; UL28 gene; UL29 gene; UL3 gene; UL30 gene; UL31 gene; UL32 gene; UL33 gene; UL34 gene; UL35 gene; UL36 gene; UL37 gene; UL38 gene; UL39 gene; UL4 gene; UL40 gene; UL41 gene; UL42 gene; UL43 gene; UL44 gene; UL45 gene; UL46 gene; UL47 gene; UL48 gene; UL49 gene; UL49A gene; UL5 gene; UL50 gene; UL51 gene; UL52 gene; UL53 gene; UL54 gene; UL55 gene; UL56 gene; UL6 gene; UL7 gene; UL8 gene; UL9 gene; UL9A gene; UL9B gene; UL9C gene; UL9D gene; UL9E gene; UL9F gene; UL9G gene; UL9H gene; UL9I gene; UL9J gene; UL9K gene; UL9L gene; UL9M gene; UL9N gene; UL9O gene; UL9P gene; UL9Q gene; UL9R gene; UL9S gene; UL9T gene; UL9U gene; UL9V gene; UL9W gene; UL9X gene; UL9Y gene; UL9Z gene; UL9AA gene; UL9AB gene; UL9AC gene; UL9AD gene; UL9AE gene; UL9AF gene; UL9AG gene; UL9AH gene; UL9AI gene; UL9AJ gene; UL9AK gene; UL9AL gene; UL9AM gene; UL9AN gene; UL9AO gene; UL9AP gene; UL9AQ gene; UL9AR gene; UL9AS gene; UL9AT gene; UL9AU gene; UL9AV gene; UL9AW gene; UL9AX gene; UL9AY gene; UL9AZ gene; UL9BA gene; UL9BB gene; UL9BC gene; UL9BD gene; UL9BE gene; UL9BF gene; UL9BG gene; UL9BH gene; UL9BI gene; UL9BJ gene; UL9BK gene; UL9BL gene; UL9BM gene; UL9BN gene; UL9BO gene; UL9BP gene; UL9BQ gene; UL9BR gene; UL9BS gene; UL9BT gene; UL9BU gene; UL9BV gene; UL9BW gene; UL9BX gene; UL9BY gene; UL9BZ gene; UL9CA gene; UL9CB gene; UL9CC gene; UL9CD gene; UL9CE gene; UL9CF gene; UL9CG gene; UL9CH gene; UL9CI gene; UL9CJ gene; UL9CK gene; UL9CL gene; UL9CM gene; UL9CN gene; UL9CO gene; UL9CP gene; UL9CQ gene; UL9CR gene; UL9CS gene; UL9CT gene; UL9CU gene; UL9CV gene; UL9CW gene; UL9CX gene; UL9CY gene; UL9CZ gene; UL9DA gene; UL9DB gene; UL9DC gene; UL9DD gene; UL9DE gene; UL9DF gene; UL9DG gene; UL9DH gene; UL9DI gene; UL9DJ gene; UL9DK gene; UL9DL gene; UL9DM gene; UL9DN gene; UL9DO gene; UL9DP gene; UL9DQ gene; UL9DR gene; UL9DS gene; UL9DT gene; UL9DU gene; UL9DV gene; UL9DW gene; UL9DX gene; UL9DY gene; UL9DZ gene; UL9EA gene; UL9EB gene; UL9EC gene; UL9ED gene; UL9EE gene; UL9EF gene; UL9EG gene; UL9EH gene; UL9EI gene; UL9EJ gene; UL9EK gene; UL9EL gene; UL9EM gene; UL9EN gene; UL9EO gene; UL9EP gene; UL9EQ gene; UL9ER gene; UL9ES gene; UL9ET gene; UL9EU gene; UL9EV gene; UL9EW gene; UL9EX gene; UL9EY gene; UL9EZ gene; UL9FA gene; UL9FB gene; UL9FC gene; UL9FD gene; UL9FE gene; UL9FF gene; UL9FG gene; UL9FH gene; UL9FI gene; UL9FJ gene; UL9FK gene; UL9FL gene; UL9FM gene; UL9FN gene; UL9FO gene; UL9FP gene; UL9FQ gene; UL9FR gene; UL9FS gene; UL9FT gene; UL9FU gene; UL9FV gene; UL9FW gene; UL9FX gene; UL9FY gene; UL9FZ gene; UL9GA gene; UL9GB gene; UL9GC gene; UL9GD gene; UL9GE gene; UL9GF gene; UL9GG gene; UL9GH gene; UL9GI gene; UL9GJ gene; UL9GK gene; UL9GL gene; UL9GM gene; UL9GN gene; UL9GO gene; UL9GP gene; UL9GQ gene; UL9GR gene; UL9GS gene; UL9GT gene; UL9GU gene; UL9GV gene; UL9GW gene; UL9GX gene; UL9GY gene; UL9GZ gene; UL9HA gene; UL9HB gene; UL9HC gene; UL9HD gene; UL9HE gene; UL9HF gene; UL9HG gene; UL9HH gene; UL9HI gene; UL9HJ gene; UL9HK gene; UL9HL gene; UL9HM gene; UL9HN gene; UL9HO gene; UL9HP gene; UL9HQ gene; UL9HR gene; UL9HS gene; UL9HT gene; UL9HU gene; UL9HV gene; UL9HW gene; UL9HX gene; UL9HY gene; UL9HZ gene; UL9IA gene; UL9IB gene; UL9IC gene; UL9ID gene; UL9IE gene; UL9IF gene; UL9IG gene; UL9IH gene; UL9IJ gene; UL9IK gene; UL9IL gene; UL9IM gene; UL9IN gene; UL9IO gene; UL9IP gene; UL9IQ gene; UL9IR gene; UL9IS gene; UL9IT gene; UL9IU gene; UL9IV gene; UL9IW gene; UL9IX gene; UL9IY gene; UL9IZ gene; UL9JA gene; UL9JB gene; UL9JC gene; UL9JD gene; UL9JE gene; UL9JF gene; UL9JG gene; UL9JH gene; UL9JI gene; UL9JJ gene; UL9JK gene; UL9JL gene; UL9JM gene; UL9JN gene; UL9JO gene; UL9JP gene; UL9JQ gene; UL9JR gene; UL9JS gene; UL9JT gene; UL9JU gene; UL9JV gene; UL9JW gene; UL9JX gene; UL9JY gene; UL9JZ gene; UL9KA gene; UL9KB gene; UL9KC gene; UL9KD gene; UL9KE gene; UL9KF gene; UL9KG gene; UL9KH gene; UL9KI gene; UL9KJ gene; UL9KK gene; UL9KL gene; UL9KM gene; UL9KN gene; UL9KO gene; UL9KP gene; UL9KQ gene; UL9KR gene; UL9KS gene; UL9KT gene; UL9KU gene; UL9KV gene; UL9KW gene; UL9KX gene; UL9KY gene; UL9KZ gene; UL9LA gene; UL9LB gene; UL9LC gene; UL9LD gene; UL9LE gene; UL9LF gene; UL9LG gene; UL9LH gene; UL9LI gene; UL9LJ gene; UL9LK gene; UL9LL gene; UL9LM gene; UL9LN gene; UL9LO gene; UL9LP gene; UL9LQ gene; UL9LR gene; UL9LS gene; UL9LT gene; UL9LU gene; UL9LV gene; UL9LW gene; UL9LX gene; UL9LY gene; UL9LZ gene; UL9MA gene; UL9MB gene; UL9MC gene; UL9MD gene; UL9ME gene; UL9MF gene; UL9MG gene; UL9MH gene; UL9MI gene; UL9MJ gene; UL9MK gene; UL9ML gene; UL9MN gene; UL9MO gene; UL9MP gene; UL9MQ gene; UL9MR gene; UL9MS gene; UL9MT gene; UL9MU gene; UL9MV gene; UL9MW gene; UL9MX gene; UL9MY gene; UL9MZ gene; UL9NA gene; UL9NB gene; UL9NC gene; UL9ND gene; UL9NE gene; UL9NF gene; UL9NG gene; UL9NH gene; UL9NI gene; UL9NJ gene; UL9NK gene; UL9NL gene; UL9NM gene; UL9NN gene; UL9NO gene; UL9NP gene; UL9NQ gene; UL9NR gene; UL9NS gene; UL9NT gene; UL9NU gene; UL9NV gene; UL9NW gene; UL9NX gene; UL9NY gene; UL9NZ gene; UL9OA gene; UL9OB gene; UL9OC gene; UL9OD gene; UL9OE gene; UL9OF gene; UL9OG gene; UL9OH gene; UL9OI gene; UL9OJ gene; UL9OK gene; UL9OL gene; UL9OM gene; UL9ON gene; UL9OO gene; UL9OP gene; UL9OQ gene; UL9OR gene; UL9OS gene; UL9OT gene; UL9OU gene; UL9OV gene; UL9OW gene; UL9OX gene; UL9OY gene; UL9OZ gene; UL9PA gene; UL9PB gene; UL9PC gene; UL9PD gene; UL9PE gene; UL9PF gene; UL9PG gene; UL9PH gene; UL9PI gene; UL9PJ gene; UL9PK gene; UL9PL gene; UL9PM gene; UL9PN gene; UL9PO gene; UL9PP gene; UL9PQ gene; UL9PR gene; UL9PS gene; UL9PT gene; UL9PU gene; UL9PV gene; UL9PW gene; UL9PX gene; UL9PY gene; UL9PZ gene; UL9QA gene; UL9QB gene; UL9QC gene; UL9QD gene; UL9QE gene; UL9QF gene; UL9QG gene; UL9QH gene; UL9QI gene; UL9QJ gene; UL9QK gene; UL9QL gene; UL9QM gene; UL9QN gene; UL9QO gene; UL9QP gene; UL9QQ gene; UL9QR gene; UL9QS gene; UL9QT gene; UL9QU gene; UL9QV gene; UL9QW gene; UL9QX gene; UL9QY gene; UL9QZ gene; UL9RA gene; UL9RB gene; UL9RC gene; UL9RD gene; UL9RE gene; UL9RF gene; UL9RG gene; UL9RH gene; UL9RI gene; UL9RJ gene; UL9RK gene; UL9RL gene; UL9RM gene; UL9RN gene; UL9RO gene; UL9RP gene; UL9RQ gene; UL9RR gene; UL9RS gene; UL9RT gene; UL9RU gene; UL9RV gene; UL9RW gene; UL9RX gene; UL9RY gene; UL9RZ gene; UL9SA gene; UL9SB gene; UL9SC gene; UL9SD gene; UL9SE gene; UL9SF gene; UL9SG gene; UL9SH gene; UL9SI gene; UL9SJ gene; UL9SK gene; UL9SL gene; UL9SM gene; UL9SN gene; UL9SO gene; UL9SP gene; UL9SQ gene; UL9SR gene; UL9SS gene; UL9ST gene; UL9SU gene; UL9SV gene; UL9SW gene; UL9SX gene; UL9SY gene; UL9SZ gene; UL9TA gene; UL9TB gene; UL9TC gene; UL9TD gene; UL9TE gene; UL9TF gene; UL9TG gene; UL9TH gene; UL9TI gene; UL9TJ gene; UL9TK gene; UL9TL gene; UL9TM gene; UL9TN gene; UL9TO gene; UL9TP gene; UL9TQ gene; UL9TR gene; UL9TS gene; UL9TT gene; UL9TU gene; UL9TV gene; UL9TW gene; UL9TX gene; UL9TY gene; UL9TZ gene; UL9UA gene; UL9UB gene; UL9UC gene; UL9UD gene; UL9UE gene; UL9UF gene; UL9UG gene; UL9UH gene; UL9UI gene; UL9UJ gene; UL9UK gene; UL9UL gene; UL9UM gene; UL9UN gene; UL9UO gene; UL9UP gene; UL9UQ gene; UL9UR gene; UL9US gene; UL9UT gene; UL9UU gene; UL9UV gene; UL9UW gene; UL9UX gene; UL9UY gene; UL9UZ gene; UL9VA gene; UL9VB gene; UL9VC gene; UL9VD gene; UL9VE gene; UL9VF gene; UL9VG gene; UL9VH gene; UL9VI gene; UL9VJ gene; UL9VK gene; UL9VL gene; UL9VM gene; UL9VN gene; UL9VO gene; UL9VP gene; UL9VQ gene; UL9VR gene; UL9VS gene; UL9VT gene; UL9VU gene; UL9VV gene; UL9VW gene; UL9VX gene; UL9VY gene; UL9VZ gene; UL9WA gene; UL9WB gene; UL9WC gene; UL9WD gene; UL9WE gene; UL9WF gene; UL9WG gene; UL9WH gene; UL9WI gene; UL9WJ gene; UL9WK gene; UL9WL gene; UL9WM gene; UL9WN gene; UL9WO gene; UL9WP gene; UL9WQ gene; UL9WR gene; UL9WS gene; UL9WT gene; UL9WU gene; UL9WV gene; UL9WW gene; UL9WX gene; UL9WY gene; UL9WZ gene; UL9XA gene; UL9XB gene; UL9XC gene; UL9XD gene; UL9XE gene; UL9XF gene; UL9XG gene; UL9XH gene; UL9XI gene; UL9XJ gene; UL9XK gene; UL9XL gene; UL9XM gene; UL9XN gene; UL9XO gene; UL9XP gene; UL9XQ gene; UL9XR gene; UL9XS gene; UL9XT gene; UL9XU gene; UL9XV gene; UL9XW gene; UL9XX gene; UL9XY gene; UL9XZ gene; UL9YA gene; UL9YB gene; UL9YC gene; UL9YD gene; UL9YE gene; UL9YF gene; UL9YG gene; UL9YH gene; UL9YI gene; UL9YJ gene; UL9YK gene; UL9YL gene; UL9YM gene; UL9YN gene; UL9YO gene; UL9YP gene; UL9YQ gene; UL9YR gene; UL9YS gene; UL9YT gene; UL9YU gene; UL9YV gene; UL9YW gene; UL9YX gene; UL9YY gene; UL9YZ gene; UL9ZA gene; UL9ZB gene; UL9ZC gene; UL9ZD gene; UL9ZE gene; UL9ZF gene; UL9ZG gene; UL9ZH gene; UL9ZI gene; UL9ZJ gene; UL9ZK gene; UL9ZL gene; UL9ZM gene; UL9ZN gene; UL9ZO gene; UL9ZP gene; UL9ZQ gene; UL9ZR gene; UL9ZS gene; UL9ZT gene; UL9ZU gene; UL9ZV gene; UL9ZW gene; UL9ZX gene; UL9ZY gene; UL9ZZ gene.

SOURCE human herpesvirus 2.  
 ORGANISM human herpesvirus 2.  
 REFERENCES  
 1. (bases 1 to 154746)  
 McGeoch, D.J., Moss, H.W., McNab, D., and Frame, M.C.  
 TITLE DNA sequence and genetic content of the HindIII 1 region in the short unique component of the herpes simplex virus type 2 genome: identification of the gene encoding glycoprotein G, and evolutionary comparisons  
 J. Gen. Virol. 68 (Pt 1), 19-38 (1987)  
 MEDLINE 87111457

```

REFERENCE          2 (bases 1 to 154746)
AUTHORS            Everett,R.D. and Fenwick,M.L.
TITLE              Comparative DNA sequence analysis of the host shutoff genes of
                  different strains of herpes simplex virus: type 2 strain HG52
JOURNAL            J. Gen. Virol. 71 (Pt 6), 1387-1390 (1990)
MEDLINE            90278430
REFERENCE          3 (bases 1 to 154746)
AUTHORS            McGeoch,D.J., Cunningham,C., McIntyre,G. and Dolan,A.
TITLE              Comparative sequence analysis of the long repeat regions and
                  adjoining parts of the long unique regions in the genomes of herpes
                  simplex viruses types 1 and 2
JOURNAL            J. Gen. Virol. 72 (Pt 12), 3057-3075 (1991)
MEDLINE            92113549
REFERENCE          4 (bases 1 to 154746)
AUTHORS            Barnett,B.C., Dolan,A., Telford,E.A., Davison,A.J. and McGeoch,D.J.
TITLE              A novel herpes simplex virus gene (UL49a) encodes a putative
                  membrane protein with counterparts in other herpesviruses
JOURNAL            J. Gen. Virol. 73 (Pt 8), 2167-2171 (1992)
MEDLINE            92356101
AUTHORS            Dolan,A.
TITLE              5 (bases 1 to 154746)
JOURNAL            Submitted (25-FEB-1997) A. Dolan, MRC Virology Unit, Church
                  Street, Glasgow, G11 5JR, UK
FEATURES
SOURCE             1. 154746
                  /organism="human herpesvirus 2"
                  /strain="HG52"
                  /db_xref="taxon:10310"
misc_feature       1 254
                  /note="a sequence"
misc_feature       235
                  /note="start of TRL (terminal copy of Long Repeat region)"
exon               440..934
                  /gene="RL1"
                  /number=1
gene               440..126810
                  /gene="RL1"
                  /join(440..934,1089..1379)
                  /gene="RL1"
                  /codon_start=1
                  /protein_id="CAB06759.1"
                  /db_xref="PID:e308397"
                  /db_xref="GI:1869821"
                  /db_xref="GI:1869821"
                  /db_xref="SWISS-PROT:P28283"
                  /translation="MSRRRCPPRRRRPPGAPVPRGAPVPRPGALPTADSQM
                  VPAVDSGTAVESAPPAASSLKRMLVPAQADSDADYAGDDLEMANSPSPESGGKAP
                  EAPHAAPAAACPPPPRRKRGPPQPLPPLPRLALRTTEYLARLSRRRPPSPAD
                  APGKACFSPRVQVRLVAMETARLARSGSMARERDRFRFRVAAAVAGPCIE
                  PEARARARARAHEDGPAEEBAAAAGSSAAAGPGRRAV"
repeat_region     446..460
                  /rpt_family="7"
repeat_region     461..547
                  /rpt_family="6"
repeat_region     943..1070
                  /rpt_family="5"
exon               1089..1379
                  /gene="RL1"
                  /number=2
                  /product="neurovirulence factor"
polyA_signal       1738..1743
                  /gene="RL1"
gene               2303..124947
                  /gene="RL2"
exon               2303..2377
                  /gene="RL2"
                  /number=1
                  /join(2303..2377,2785..3462,3644..5368)
                  /gene="RL2"
                  /codon_start=1
                  /protein_id="CAB06760.1"

                  /db_xref="PID:e304156"
                  /db_xref="PID:1869822"
                  /db_xref="GI:1869822"
                  /db_xref="SWISS-PROT:P28284"
                  /translation="MEPRPGTSSRADGPPRRPPTGCTQPAAPHANGMINDMOWAS
                  SDSEEFVEIGSDDDLHSDTSSEKSGIDTEFEGLMDAATPPAPRAERKSGSTPAD
                  AQSGGGGPGVEEAEAGGGDCAVCTDITAPLRQAQSFCLHPCFIPCKWTIPR
                  NTPCLNTPVAYLIVGTASGSEFTIPVDPPRTVRAEAAVAGTAVIDITMGNTPT
                  APRSLSGHTVRLALSPTPMPGTDIDDDLDADYDVPAPRAPRRGGGAGATRGCT
                  SOPATAPAPGAPRRSSSGGAPLRAVGSGSGGPAVAAPVPAVSLPAGGRQ
                  ARVGENAAAGRTPPAPRAQEPPIYISDPPSPRRRAGPGLSTVSSSAOY
                  SSGGGGGLUOSSGRRAPRAVAFRVRSPPRAAAVVASADAAPAPAPVVDH
                  RAPRSRTQATDITQAQSLGRAGATDARGSGGPPAEGGPVPRTNTPGAAPAAEBA
                  AAPRRKRGSDDGPPAASSASASAPRSPPLAPOGVAKRAAPRDPDSDRGHPL
                  APAGAAPPASPSOAAVAAASASSASASSASASSASASSASASSASASSAS
                  SSSAGGSGSVASASGGERRENTIGPRAAPRQRCARKTRAEGPPEGARDPAP
                  GTRVLYIAVSSVYALAPYVNTKYTDCTLPVDMETGILGATVVLVDGTGNADLIR
                  AAPAMSRKLLPPEHANCVRPPDPPPPASSEWNSLMTVGNMLFDGCTLVGLDPRH
                  GLRRHFWRSREOGAPAPADAPAGHG"
exon               2785..3462
                  /gene="RL2"
                  /number=2
repeat_region     3468..3607
                  /rpt_family="4"
exon               3644..5368
                  /gene="RL2"
                  /number=3
                  /function="modulator of cell state and gene expression"
repeat_region     4655..4771
                  /rpt_family="3"
misc_feature       complement(4811)
                  /gene="RL2"
polyA_signal       5618..5623
                  /note="proposed LAT splice acceptor site"
misc_feature       complement(7025)
                  /gene="RL2"
                  /note="proposed LAT splice donor site (5' end of stable
                  LAT species)"
misc_feature       complement(7732)
                  /gene="RL2"
                  /note="primary transcription initiation site for latency
                  associated transcripts (LAT's)"
repeat_region     9047..9193
                  /rpt_family="2"
repeat_region     9276..9297
                  /rpt_family="1"
misc_feature       9298
                  /gene="RL2"
gene               9427..10101
                  /note="start of UL (Long Unique region)"
                  /gene="UL1"
                  /codon_start=1
                  /product="viral glycoprotein L"
                  /protein_id="CAB06761.1"
                  /db_xref="PID:e304262"
                  /db_xref="PID:1869823"
                  /db_xref="GI:1869823"
                  /db_xref="SWISS-PROT:P28278"
                  /translation="MGFVCLFGLVWVGAWGAGSGQATPEVLRSVAKRVEGDLRVC
                  MRPPADDVSMRKEAPVIDYKARIDGFLFRLHPGGLDTPLMDHQAQRAIVNPLFRAQ
                  FIEDSHSVFPADTOETTRKALYKEIRDALSKRQOAVSHAVRAGCVNFDYSRRRC
                  VGRDLRPAINTSTWEPVSDDEASSOSKPLATOPVYALSNAPRRVSPTRGRRRH
                  TRLRN"
gene               10211..10978
                  /gene="UL2"
gene               10211..10978
                  /codon_start=1
                  /product="uracil-DNA glycosylase"
                  /protein_id="CAB06762.1"

```



/number=2  
/product="neurovirulence factor"  
1738. 1743  
/gene="RL1"  
2303. 124947  
/gene="RL2"  
2303. 2377  
/gene="RL2"  
/number=1  
Join(2303. 2377, 2785. 3462, 3644. 5368)  
/gene="RL2"  
/codon\_start=1  
/protein\_id="CAB06760.1"  
/db\_xref="PID:e304156"  
/db\_xref="PID:g1869822"  
/db\_xref="GI:1869822"  
/db\_xref="SWISS-PROT:P28284"  
/translation="MEPRGTSRADPGEPPEPPROTPTGTPAPAPHAMGLNMMOMLAS  
SDSEETEVGISDDDLHRDSTSEAGTDEMEFAGIMDAATPAPAPAROCSSPTPAD  
AGSCGGGPGVEEAEAGGCGVCACTEETIAPPLRCOSFPLHPFCLCMKWTIPLR  
NTCPICNTVAITLIVGTVASGSFSTIPVNDPRTVEAAVAAGTADFIWTGNPRT  
APRSJLGHITVRLSPTPMWTGTDDEDDLDVDPVPPAPAPAPAGGAGATGT  
SOPATRPAPGAPSSSSGAPLRAGVSGSGGGAVAAYPRVAPSLPPAGGGAGT  
ARRVEDAAAEGRTPPAROPPAOEPPIVISDPPSPRRAPGEPISFVSSSSAOV  
SSGPGGGGLPOSSGRRAPRAVAPPRVPPRAAPVYASADAAGPAPVPPVDAH  
RAPRRMTOAQDTQAOISIGRAGTADARSGGPGAEAGGTVPRGTVTPGAHAEGA  
NAPRRKROSDGPPAASSSSAAAPRSPLAQGVAKRAAPRAAPDSGGRGAPL  
APASGAPPSAPSSQAANVAASSSSASSSSASSSSASSSSASSSSASSSSA  
SSSAGAGSVASAGAGERRSLGAPRAAPGAPKCAKTRHAGGEPGAPGADP  
GLTRPLIAGVSVVLAIVKTYGTGDLPLTDMETGHIGAYVLYNQTGVALLR  
AAAPASRRTLLPEHARNCVRPPDYPTPASEMSLMTVGNMLFDOSTLVGALDFH  
GIRSRHPSREGAPAPAGDAPAGHE"  
2785. 3462  
/gene="RL2"  
/number=2  
3468. 3607  
/rpt\_family="4"  
3644. 5368  
/gene="RL2"  
/number=3  
/function="modulator of cell state and gene expression"  
4655. 4771  
/rpt\_family="3"  
complement(4811)  
/gene="RL2"  
/note="proposed LAT splice acceptor site"  
5618. 5623  
/gene="RL2"  
complement(7025)  
/gene="RL2"  
/note="proposed LAT splice donor site (5' end of stable  
LAT species)"  
complement(7732)  
/gene="RL2"  
/note="primary transcription initiation site for latency  
associated transcripts (LAT's)"  
9047. 9193  
/rpt\_family="2"  
9276. 9297  
/rpt\_family="1"  
9298  
/gene="RL2"  
/note="start of UL (long unique region)"  
9427. 10101  
/gene="UL1"  
9427. 10101  
/gene="UL1"  
/codon\_start=1  
/product="viral glycoprotein I"  
/protein\_id="CAB06761.1"  
/db\_xref="PID:e304262"  
/db\_xref="PID:g1869823"  
/db\_xref="GI:1869823"

/db\_xref="SWISS-PROT:P28278"  
/translation="MGFVCLFGLVVMGANGGSGQATEYLRVSIYAKNEVSDILRYP  
KRTPADVMSREAPSVIDYARIDGIFLYHCPGIDTFEMDHAARAYLVNPELPAG  
FLEDLSHVSFPADTOETTRALYKEIRALGSRKROAVSHAPVAGCVDFYSRPRC  
VGRDLRPANNTSTWEPVPSDDEASSGSKPLATGPVYALSNAPRRARVPRGRRH  
TRLRN"  
10211. 10978  
/gene="UL2"  
10211. 10978  
/gene="UL2"  
/codon\_start=1  
/product="uracil-DNA glycosylase"  
/protein\_id="CAB06762.1"  
/db\_xref="PID:e304157"  
/db\_xref="PID:g1869824"  
/db\_xref="GI:1869824"  
/db\_xref="SWISS-PROT:P28275"  
/translation="NFSASTPEQPLGSGATPPLPVSLDMAAFRRATLIDAMR  
FULEELANPLTARILAEYDRRCQTEEVLPREDVSWTRCTPDVVRVITIGDPYH  
HFGQAHGLAFSVRADVPVPSLRVTLAAVKNCYPPARSGCLEKMARADVILLNTT  
LTVKRGAAASHKLGMDREFGVGVORLAARGLPMLGAAQAIVARIPDRQHYVLK  
FSHPSPLSKVPKQGHFLAANRYETDIMPDMVS"  
11019. 11024  
/gene="RL2"  
11033. 11734  
/gene="UL3"  
11033. 11734  
/gene="UL3"  
11033. 11734  
Query Match 13.8%; Score 63; DB 17; Length 154746;  
Best Local Similarity 48.7%; Pred. No. 0.092;  
Matches 171; Conservative 0; Mismatches 180; Indels 0; Gaps 0;  
polyA\_signal  
gene  
CDS  
37 GCCCTGCGCCGCGCAGAGCGCGAGAGCCGCGTCCGCTGAGAGTGGGGTGA 96  
11  
Db 148472 GCCGGGGGAGACGGCGCGGGGAGACGGCGCGGGGAGACGGCGCGGGG 148413  
11  
QY 97 GGTGCGCGGAGCGTACCGACGCGCGGGAGTACCTCCGCTCACACCTCGGGCTGCATTTC 156  
11  
Db 148412 GACGGGCGCGGGGAGCGCGCGGGGAGCGGGCGGGGAGCGGGCGGGGCGGGG 148353  
11  
QY 157 CCGTGGTGGCGCGCGAGACGCTGGCGCGGGCGGGGCGGGGCGGGGCGGGGCGGC 216  
11  
Db 148352 GCGGGGGGCGCGGGGGCGCGGGGGGAGCGGGGGGAGCGGGGGGAGCGGGGGGAGC 148293  
11  
QY 217 GCGGCGGCGCTGCTGACAGTGAAGAGCGCGGGGCGGGGCGGGGCGGGGCGGGC 276  
11  
Db 148292 GCGGCGGCGGGGAGCGGGGGGAGCGGGGGGAGCGGGGGGAGCGGGGGGAGCGGGG 148233  
11  
QY 277 CGGCTCCTTAAAGCGCGGAGCGCGGAGCGGAGGTGCTCTGTGCGCGCAGCGCGAGCG 336  
11  
Db 148232 GACGGGGGAGACGGGGGAGCGGGGGGAGCGGGGGGAGCGGGGGGAGCGGGGGGAGC 148173  
11  
QY 337 CCGGCGGAGACCGCGAGCGTGAAGCGCGCGCGCTGCTGCACTCCGGGAGC 387  
11  
Db 148172 CCGGGGAGCGGGGGGAGCGGGCGCGGGGGGAGCGGGGGGAGCGGGGGGAGC 148122  
11  
RESULT 12  
LOCUS HSU40583 1977 bp mRNA PRI 19-DEC-1995  
DEFINITION Human alpha 7 neuronal nicotinic acetylcholine receptor mRNA,  
complete cds.  
ACCESSION U40583  
NID g1125076  
VERSION U40583.1 GI:1125076  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homini; Homo.  
REFERENCE 1 (bases 1 to 1977)  
AUTHORS Logel,J., Dreding,C., Barnhart,M., Antle,C. and Leonard,S.  
TITLE Nucleotide Sequence and Transcript Size of the Alpha-7 Neuronal

JOURNAL  
REFERENCE  
2 (bases 1 to 1977)  
AUTHORS  
Leonard, S.  
TITLE  
Submitted (13-NOV-1995) Sherry Leonard, University of Colorado  
JOURNAL  
Health Sciences Center, C-268-71 Pharmacology, 4200 E. Ninth Ave,  
Denver, CO 80262, USA

FEATURES  
source  
Location/Qualifiers  
1. .1977  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="pBSshalpa7"  
/clone\_1db="Clontech lambda gt10 cDNA library: human  
hippocampus"  
/sex="male"  
/tissue\_type="hippocampus"  
/dev\_stage="20 yr old adult"  
1. .7  
8. .73  
8. .1516  
/codon\_start=1  
/product="alpha 7 neuronal nicotinic acetylcholine  
receptor"  
/protein\_id="AA83561.1"  
/db\_xref="PID:g1125077"  
/translation="MRCSPGVWLAALSLHVSLOGEQRKLYKELVKNYPLERPV  
ANDSOPILVYSNLQIMDVDERKNVLTINILQMSWMDHYLFGKSSYIDVPEP  
PDQIMKPDILLYNSADEREDAFENHNLVNPSCQYLPBFGKSSYIDVPEP  
VOHCKLFGSGSYGMSLDLQMOEADISGYPNGEMDLYIPKREBERECCKEEP  
DNYETVMRRRTLYGKLNILIRPVILSALALYFLPADGEGISGITYLSITFM  
LVAEIRPRTSDSPLTAYQFASTMITVGLSVVYVYALDYTHHDPDGKMKWTIVL  
LWMCWFLMKRPRGKEDKVRACQKQRCSLASVESAAPPASNGMLYIGFGL  
DVHCVPTPDGYYVCGRMASPTNDEHILHGQPPEDDPLAILEVRITANRFRCQ  
DESEAVCEWKFACVVDRLCLMAFVFTICTIGILMSAPNEVEAVSKQEA"  
1514. .1977

3'UTR  
polya\_site  
BASE COUNT 426 a 567 c 524 g 460 t  
ORIGIN

Query Match 13.6% Score 62; DB 10; Length 1977;  
Best Local Similarity 100.0%; Pred. NO. 0.54; Indels 0; Gaps 0;  
Matches 62; Conservative 0; Mismatches 0;

386 ACTCAACATGGCTGCTGCCGCGGAGGCGTGTGCGCTGCGCGCTGCTGCA 445  
|||||  
Db 1 ACTCAACATGGCTGCTGCCGCGGAGGCGTGTGCGCTGCGCGCTGCTGCA 60

QY 446 CG 447  
||  
Db 61 CG 62

RESULT 13  
LOCUS SCE94 38532 bp DNA BCT 12-APR-1999  
DEFINITION Streptomyces coelicolor cosmid E94.  
ACCESSION AL049628  
VERSION 94585581  
KEYWORDS  
A1049628.1 GI:4585581  
acetyltransferase; ATP-binding; Clp-family ATP-binding protease;  
cseB; cseC; DNA repair; DNA-binding; ECF sigma factor; exporter;  
L-aspartate oxidase; lipA; lsr2; mltv; nadB; nadC;  
nicotinate-nucleotide pyrophosphorylase; pantoate-amino acid ligase;  
putative adenine glycosylase; response regulator; sensory histidine  
kinase; sigE.  
Streptomyces coelicolor.  
Streptomyces coelicolor.  
Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;  
Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
REFERENCE  
1 (bases 1 to 38532)

AUTHORS  
Redenbach, M., Kieser, H.M., Denaplatte, D., Eichner, A., Cullum, J.,  
Kinsahl, H. and Hopwood, D.A.  
TITLE  
A set of ordered cosmids and a detailed genetic and physical map  
for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
Mol. Microbiol. 21 (1), 77-96 (1996)  
JOURNAL  
MEDLINE  
REFERENCE  
97000351  
AUTHORS  
2 (bases 1 to 38532)  
Oliver, K. and Harris, D.  
JOURNAL  
Unpublished  
REFERENCE  
3 (bases 1 to 38532)  
Bentley, S.D., Parkhill, J., Barrell, B.G. and Rajandream, M.A.  
AUTHORS  
Direct Submission  
JOURNAL  
Submitted (12-APR-1999) Streptomyces coelicolor sequencing project,  
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.  
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,  
Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT  
Notes:  
Streptomyces coelicolor sequencing at The Sanger Centre is funded  
by the BBSRC.  
Details of S. coelicolor sequencing at the Sanger Centre are  
available on the World Wide Web.  
(URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>)  
CDS are numbered using the following system eg SC787.01c. SC (S.  
coelicolor), 787 (cosmid name), .01 (first CDS), c (complementary  
strand).  
The more significant matches with motifs in the PROSITE database  
are also included but some of these may be fortuitous.  
The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for  
CDS which show significant similarity to other CDS in the database.  
The position of possible ribosome binding site sequences are given  
where these have been used to deduce the initiation codon.  
Gene prediction is based on positional base preference in codons  
using a specially developed Hidden Markov Model (Krogh et al.,  
Nucleic Acids Research, 22(22):4768-4781(1994)) and the FramePlot  
program of Bibb et al., Gene 30:157-66(1984) as implemented at  
<http://www.nhn.gov.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the  
correct initiation codon. Where possible we choose an initiation  
codon (atg, gtg, ttg or (att)) which is preceded by an upstream  
ribosome binding site sequence (optimally 5-13bp before the  
initiation codon). If this cannot be identified we choose the most  
upstream initiation codon.  
IMPORTANT: This sequence MAY NOT be the entire insert of the  
sequenced clone. It may be shorter because we only sequence  
overlapping sections once, or longer, because we arrange for a  
small overlap between neighbouring submissions.  
Cosmid E94 lies between E7 and E126 on the Aser-E genomic  
restriction fragment.

FEATURES  
source  
Location/Qualifiers  
1. .38532  
/organism="Streptomyces coelicolor"  
/strain="A3(2)"  
/db\_xref="taxon:1902"  
/clone="cosmid E94"  
complement(1..1512)  
/gene="SCE94.01c"  
1. .105  
/gene="SCE94.01c"  
/note="Nominal overlap with Streptomyces coelicolor E7."  
complement(1..1512)  
/gene="SCE94.01c"  
/note="SCE94.01c, partial CDS unknown, alanine-rich  
protein, len: >503aa"  
/codon\_start=1  
/transl\_table=11  
/label="SCE94.01c  
/product="alanine-rich hypothetical protein"  
/protein\_id="CAB40850.1"  
/db\_xref="PID:e1424928"  
/db\_xref="PID:g4585582"  
/db\_xref="GI:4585582"

[illegible]

```

/codon_start=1
/transl_table=1
/label_SCE94.04c
/product-"hypothetical protein"
/protein_id="CAB40853.1"
/db_xref="PID:e1424931"
/db_xref="PID:g14585585"
/db_xref="GI:4585586"
/translation-"MKFSLKSVNSYSFGQTAFELTANTSGDENVLGRHAF
ITPASGDAYMAWSDDCVEAGSILRYVTGSGITTYVKWDRKPSAECCTPAGASK
GTALVKAAGFAGEKVOTSEFLNKD"
5169..5573
/gene="SCE94.05"
/gene=".5573
5169..5573
/gene="SCE94.05"
/note="SCE94.05, unknown, proline-rich, len: aa"
/codon_start=1
/transl_table=1
/product-"hypothetical proline-rich protein"
/protein_id="CAB40854.1"
/db_xref="PID:e1424932"
/db_xref="PID:g14585586"
/db_xref="GI:4585586"
/translation-"MOLGRIDPDAGSASSAPFYVTAAPPPLADPADPPDSPPDFE
EPRLAPREPPPEPEPLBSPLDSSRRPGMALVGPDEBPVIDGTFPLAPSSHS
PIPPSPPETTLTHVINSANRAKRDSSTLRQ"
5816..6742
/gene="SCE94.06"
/gene=".6742
5816..6742
/gene="SCE94.06"
/note="SCE94.06, muty, putative adenine glycosylase, len:
308aa; previously partially sequenced therefore partially
identical to TR:E1358524 (EMBL:AJ31213) muty, putative
adenine glycosylase from Streptomyces coelicolor (183 aa)
fasta scores: opt: 1220, z-score: 1357.4, E(): 0, (100.0%
identity in 183 aa overlap). Also similar to many others
eg. SW:MUTY-SMUTY muty, adenine glycosylase from

Query Match      13.5%; Score 61.6; DB 1; Length 38532;
Best Local Similarity 47.0%; Pred. No. 0.23;
Matches 190; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY    26  GGCCCTGGGCACGCCCCCTGGCCCGGCCAAGAGGGCGGAGGCCCGCTCGGTGGAGA  85
      |||||
Db    33300  GGTCCTCACGAATCACCCGGATTAAAGCCCATGAGAAAGAGCGGCATGACGGCGGGGGCAG  33359
QY    86  CTGGGGGTTGAGAGTGCCCGAGAGCTAACCCAGAGCGCGGGAGTAGACCCCGCTCAACACTCG  145
Db    33360  GAGCGCGGAGAGTAGTGTCATAGAGACAGGGTAGCCAGAGCCCTTCACAGCGCGGCGCAGG  33419
QY    146  GGCTCAGTTCCTTGCTGGTGGCCCGAGACGCTGGCCCGGGCTGAGAGATGAGCGGGCGC  205
Db    33420  TGTGTCCCGGAGATCCGGGGGCGGGGCGCGGCGGGCGGCGCTTGCGGGGGGGGAAACACC  33479
QY    206  GGGAGCGGGGCGGGGGGCTCGTACGTTGAGAGAGCGCGCGGGGCGGGGCGGGCGGCGG  265
Db    33480  TGGCCGGGGGTTGGGAGACGGGCGCTTTGGGGGTTCCGGGCGCGGGCGCGGGGCGGGGAC  33539
QY    266  GGGCGCGCGCCCGGCTCCTTAAAGCGCGCGAGCGCGAGCGGCGAGAGTGCCTGTGGGCGG  325
Db    33540  GGTTCGCGGGGCGGCTCTGTGGGGGCTTCCTTGCAGACATGCCGGCGCGCGGGGACGCGG  33599
QY    326  CAGGGCGAGGCGCGGGGCGACAGCCGAACGTGAGAGCGCGCGCGCTCGTCACTCCGGG  385
Db    33600  TTGAGGCGGGGTGCGGGGCGGGGCGCGGGGAGGGGAGCGGTGAGAGCGGTGGCGCGGTTG  33659
QY    386  ACTCAACATGCGCTGCTCGCCGAGGAGCGCTGTGAGGCTAGGCGCTGG  429
Db    33660  GGCCTCTCGGGGGGCGAGAGCGGGCCAGAGAGGCGCGGCTGCGGCGGG  33703

```



```

CH19R30879/c CH19R30879 41936 bp DNA Pri 01-APR-1997
LOCUS Homo sapiens DNA from chromosome 19-cosmid R30879 containing USF2,
DEFINITION genomic sequence.
ACCESSION AD000684
AD000684
VERSION 1
AD000684.1 GI:1905917
KEYWORDS chromosome 19; transcription factor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Euhelina; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 41936)
REFERENCE
AUTHORS Lamerdin,J.E.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1996) J.E. Lamerdin, Human Genome Center,
Lawrence Livermore National Laboratory, 7000 East Ave, Livermore,
CA, USA, 94551 janelacgt.llnl.gov owetornak.llnl.gov
GDB:S:1010600.
COMMENT Human Genome Center
Biology and Biotechnology Research Program
Lawrence Livermore National Laboratory
7000 East Avenue
Livermore, CA 94550 USA
constructed at LLNL from flow-sorted chromosomes
from hybrid 5HL2-B, which carries chromosome 19 as its only human
chromosome.
FEATURES
source Location/Qualifiers
1. 41936
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone_lib="R30879"
/clone_type="fibroblast"
/clone_lib="LL19NC03 R chromosome 19-specific cosmid
library"
repeat_region complement(1..34)
/note="repeat match = HSA115306; putative"
/rpt_family="Alu"
repeat_region complement(811..1099)
/note="repeat match = HSA100208; putative"
/rpt_family="Alu"
repeat_region 1335..1479
/note="repeat match = HSA105512; putative"
/rpt_family="Alu"
repeat_region 1367..2305
/note="repeat match = HSA102043; putative"
/rpt_family="Alu"
repeat_region complement(3093..3356)
/note="repeat match = HSA102254; putative"
/rpt_family="Alu"
repeat_region 4030..4168
/note="repeat match = HSA103448; putative"
/rpt_family="Alu"
repeat_region complement(4031..4288)
/note="repeat match = HSA101019; putative"
/rpt_family="Alu"
repeat_region 4435..4533
/note="repeat match = HSA102522; putative"
/rpt_family="Alu"
repeat_region 4448..4565
/note="repeat match = HSA101715; putative"
/rpt_family="Alu"
repeat_region 4926..5218
/note="repeat match = HSA106523; putative"
/rpt_family="Alu"
repeat_region complement(5804..6091)
/note="repeat match = HSA104878; putative"
/rpt_family="Alu"
repeat_region complement(6051..6102)
/note="repeat match = HSA104290; putative"
/rpt_family="Alu"
repeat_region 7131..7434
/note="repeat match = HSA102525; putative"
/rpt_family="Alu"
repeat_region 7137..7589
/note="repeat match = HSA100829; putative"
/rpt_family="Alu"
gene 7911..16824
/feature="LISCH7"
CDS 15802..15933,16075..16700,16790..16824
/feature="LISCH7"
/function="human homolog of Mus musculus B6CBA Lisch7
mRNA"
/note="putative"
/codon_start=2
/product="liver-specific bHLH-Zip transcription factor"
/protein_id="A051178.1"
/db_xref="pid:91905918"
/db_xref="GI:1905918"
/translation="ADLTEDOTAMGDSGVYGSVSNODLQGNNAEYAEILVGRISG
VAELLPEFGQPIEGWLFYVYVCLAEFLITLLIGMCCPHTCCCVGPGCCPDG
CCEPARYAAGKAATSGVPSIYAPSTIAHLSPATPPRPAIRPGAPANGPGGYPGV
DRSSGVSQGRIOASQODSMKLYMERELANFDESRPGSPGARGVREKMSYTSB
EDMRSRPSRGPALTPIDEMGSHSPSRPGMDPEAPRQAGGVARPRKASVDA
LDLTPSTAESGSRSPSTNGSRRAVMPRPSRDLDYQDDSDRPPSRPYDDE
RSRRRPADPRSHHRRDRDRDNGSRGDLPYDGRLLAEAVRKKGSEERRRPHKEEE
EAYPPAPPPYSETDSQASRERLRKDLASRESLYV"
8303..8386
/note="repeat match = HSA105819; putative"
/rpt_family="Alu"
8325..8609
/note="repeat match = Alu; putative"
/rpt_family="Alu"
repeat_region complement(9257..9546)
/note="repeat match = HSA105084; putative"
/rpt_family="Alu"
repeat_region complement(9513..9555)
/note="repeat match = HSA100054; putative"
/rpt_family="Alu"
repeat_region 9599..9759
/note="repeat match = HSA117230; putative"
/rpt_family="Alu"
repeat_region 9639..9813
/note="repeat match = HSA106130; putative"
/rpt_family="Alu"
repeat_region 10359..10498
/note="repeat match = HSA106554; putative"
/rpt_family="Alu"
repeat_region 11510..11657
/feature="LISCH7"
/gene="LISCH7"
/note="similarity: gb|H42128|H42128: y061b12.r1 Homo
sapiens cDNA clone 18239 5'; putative"
11957..12072
/note="repeat match = HSA10157; putative"
/rpt_family="Alu"
repeat_region 11981..12253
/note="repeat match = Alu; putative"
/rpt_family="Alu"
repeat_region 12168..12279
/note="repeat match = HSA103934; putative"
/rpt_family="Alu"
repeat_region complement(12334..12610)
/note="repeat match = HSA100279; putative"
/rpt_family="Alu"
repeat_region complement(12391..12982)
/note="repeat match = HSA102441; putative"
/rpt_family="Alu"
repeat_region complement(12645..13088)
/note="repeat match = HSA102504; putative"
/rpt_family="Alu"
repeat_region 16521..16586
/gene="LISCH7"
/note="similarity: gb|R39945|R39945: yf52e09.s1 Homo

```

Query Match	Best Local Similarity	Score	DB	Length	Indels	Gaps
Matches 155; Conservative 0; Mismatches 134; Indels 3; Gaps 1;	53.1%;	61.6;	10;	41936;		
55	GGCGGAGAGCCGAGAGCCCGCGTGTGAGAGATGGGGGTGAGAGTGGCCGAGGCTACCC	114				
Db	18136	GGCGGAGAGCCGAGAGCCCGCGTGTGAGAGATGGGGGTGAGAGTGGCCGAGGCTACCC	18127			
QY	115	AGCGCGGAGAGTACTCCCGCTACACCTCGGGGCTGCAAGTTCCCGGGTGGCCGCGGAGA	174			
Db	18126	GGCGGAGAGAGAGCGGTGCGGAGCGAGCGGATCCAGACCC---GGGTCAGATGTC	18070			
QY	175	CGCTGGCCCGGAGCTGGAAGATGGCGGGGCGGAGACGAGGGGCGGGGCTGCTGAC	234			
Db	18069	CATGGGGGGGCGGGGCGGGGGGCGCGGGGCGCGGGGCGGGGAGAGGAGAGGAG	18010			
QY	235	GTGAGAGAGCGCGCGGGGCGGGGCGGGGCGCGCGCGCGCGGCTCTTAAAGCGCG	294			

Db	18009	GAGGAGGAGGGAGAGGGCGGGCCGCCGAGGAGCCCCCGAGCCGGCGCTCACGCCGCGG	19950
OY	295	CGAGCCGAGCGGCGAGGTGCCTGTGTGGCCGACAGGCGCAGCCCGGCGACA	346
Db	17949	GCAGGAGGCGAGACGAGATACGGGAGCGCGTCCGCGTGATCAGGGGACA	17898

  

<b>RESULT 15</b>			
HUSUF2/c	HUSUF2	14440 bp	PRI
LOCUS	DEFINITION	H.sapiens USF2 gene.	26-JUN-1997
ACCESSION	NID	GI806093	
VERSION	KEYWORDS	X07661.1 GI:1806093	
SOURCE	SOURCE	human.	
ORGANISM	ORGANISM	Homo sapiens	
REFERENCE	REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;	
AUTHORS	AUTHORS	Primates; Catarrhini; Homnidae; Homo.	
TITLE	TITLE	1 (bases 1 to 14440)	
JOURNAL	JOURNAL	Groenen,P.M., Garcia,E., Debeer,P., Devriendt,K., Frys,J.P. and	
MEDLINE	MEDLINE	Van de Ven,W.J.	
REFERENCE	REFERENCE	Structure, sequence, and chromosome 19 localization of human USF2	
AUTHORS	AUTHORS	and its rearrangement in a patient with multicystic renal dysplasia	
JOURNAL	JOURNAL	Genomics 38 (2), 141-148 (1996)	
TITLE	TITLE	2 (bases 1 to 14440)	
COMMENT	COMMENT	Groenen,P.	
FEATURES	FEATURES	Direct submission	
SOURCE	SOURCE	Submitted (27-AUG-1996) P. Groenen, Center for Human Genetics,	
		Laboratory for Molecular Oncology, Herestraat 49, 3000 Leuven,	
		BELGIUM	
		Related sequence: X90823-X90826.	
		Location/Qualifiers	
		1..14440	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/chromosome="19"	
		/haplotype="2n"	
		/map="q13.1"	
		/clone_lib="LNL 7058"	
		/dev_stage="adult"	
		join(3688..3749,4031..4077,4166..4284,4388..4588,	
		5032..5182,5303..5390,5668..5726,13284..13378,	
		13332..13660,13753..14380)	
		/gene="USF2"	
		3688..14380	
		/gene="USF2"	
		3688..3749	
		/gene="USF2"	
		number=1	
		join(3688..3749,4031..4077,4166..4284,4388..4588,	
		5032..5182,5303..5390,5668..5726,13284..13378,	
		13532..13660,13753..13842)	
		/gene="USF2"	
		/codon_start=1	
		/protein_id="CAA68942.1"	
		/db_xref="PIR:E291919"	
		/db_xref="PIR:G1806094"	
		/db_xref="GI:1806094"	
		/db_xref="SWISS-PROT:O15853"	
		/translation="MDMLDPLGDLPASATAAASHDKPEAEVEIIEQSGDGAE	
		ETLAVALTSVOAAARGNDINIOYRTFTNGSQVTRVVQVIDGLDGGDTAGAVSV	
		VSAARAGCAQTGVGVDAORPGPAPASVPFGPAAPFLAVIONPSNGSPAAEA	
		VSGEKRAIVPASVSGDTTAVSVOTDSLAGGFYMMTPQVDLTGRTIAPRT	
		HPYSKIGRTIPDERRRRAQHNEVERRDKNINMWIOLSKTIIPDCNADNSKGASK	
		GISLSKADYIRELEHOTNRMQEFTKEKERLOMNNELLROOIEELKENNALIRLOLO	
		NHLEWVGEGTRQ"	
		3750..4030	
		/gene="USF2"	
		/number=1	
		4031..4077	

  

CDS	exon	gene	mRNA	intron	exon
-----	------	------	------	--------	------

```

Intron      /gene="USF2"
             /number=2
             4078..4165
             /gene="USF2"
             /number=2
             4166..4284
             /gene="USF2"
             /number=3
             4285..4387
             /gene="USF2"
             /number=3
             4388..4588
             /gene="USF2"
             /number=4
             4589..5031
             /gene="USF2"
             /number=4
             5032..5182
             /gene="USF2"
             /number=5
             5183..5302
             /gene="USF2"
             /number=5
             5303..5390
             /gene="USF2"
             /number=6
             5391..5667
             /gene="USF2"
             /number=6
             5668..5726
             /gene="USF2"
             /number=7
             5727..13283
             /gene="USF2"
             /number=7
             13284..13378
             /gene="USF2"
             /number=8
             13379..13531
             /gene="USF2"
             /number=8
             13532..13660
             /gene="USF2"
             /number=9
             13661..13752
             /gene="USF2"
             /number=9
             13753..14380
             /gene="USF2"
             /number=10
             /number=10

BASE COUNT  2766 a 4210 c 4353 g 3110 t      1 others
ORIGIN

```

Query Match 13.5%; Score 61.6; DB 10; Length 14440;  
 Best Local Similarity 53.1%; Pred. No. 0.32;  
 Matches 155; Conservative 0; Mismatches 134; Indels 3; Gaps 1;

```

OY  55  GCGCGAGGCGCGAGAGCCCGCTGCTGAGACTGGGGGTGAGAGTCCCGGAGCGTACCC 114
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  3807 GCGCGGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGGATCTTACT 3748

OY  115  AGCGCGCGGAGTACCTCCCTCAACACTGCGGCTGCAAGTCCCTGGGTGCCCGCCGAGA 174
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  3747 GCGCGCGCGAGCGAGCGGTGCGCGAGCGAGCGGATCCAGACC---GGTCCAGCATGTC 3691

OY  175  CGCTGGCCCGCGGCTGAGAGATGCGGCGCGGCGGCGGCGGCGGCGGCGGCGCTGCTAC 234
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  3690 CATGGGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAGGAGGAG 3631

OY  235  GTGAGAGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTTAAAGCGCG 294
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  3630 GGAGGAGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGCTACGCGCGCG 3571

```

```

OY  295  CGAGCCGAGCGCGAGGTGCTCTGTGCGCCGAGGCCGAGGCCCGGCGGAGCA 346
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  3570 GCAGAGGAGGAGCGGAGAGATACGAGAGCCGCTGCGCTGATCACGGGAGCA 3519

```

Search completed: September 17, 1999, 22:02:31  
 Job time: 16334 sec

Sun Sep 19 10:27:05 1999

us-08-956-518a-94.rng

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:33:39 ; Search time 425.19 seconds

(without alignments)  
268.910 Million cell updates/sec

Title: US-08-956-518a-94

Perfect score: 457

Sequence: 1 AAGACGCAAGGAGAGAGTAC.....CTCTGACGATTAAGCCAC 457

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database: N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	127	27.8	1876	1	T48239	Neuronal nicotinic
2	127	27.8	1876	1	V12197	Human neuronal nic
3	69.2	15.1	3861	1	Q51543	Lipopolysaccharide
4	68.8	15.1	114955	1	X53491	Human adenosine A1
5	65.2	14.3	114955	1	X53491	Human adenosine A1
6	64	14.0	3198	1	X02874	Human IL-1ra BAC C
7	60.6	13.3	1281	1	Q23296	HSV-1 (MGH-10) ICP
8	57.8	12.6	11820	1	V18130	Human chromosome 1
9	56.4	12.3	12001	1	Q76213	HSV L/ST region. H
10	56	12.3	1120	1	V60339	CDNA sequence of f
11	53	11.6	4257	1	V10362	Infected cell prot
12	53	11.6	4257	1	V68820	The nucleotide seq
13	52.4	11.5	12001	1	Q76213	HSV L/ST region. H
14	52.4	11.5	5228	1	V81384	Nucleic acid enco
15	51.8	11.3	4403	1	Q47927	Paired basic amino
16	51.8	11.3	2218	1	Q47929	Human enzyme-relat
17	51	11.2	6225	1	X55273	Human adeno-associ
18	50.8	11.1	407	1	V44430	Mycobacterium tube
19	50.8	11.1	407	1	V64539	M. tuberculosis im
20	50.4	11.0	7193	1	V50431	Streptomyces clav
21	50.2	11.0	1335	1	Q23295	HSV-1 (F) ICP34.5
22	50	10.9	4067	1	Q63192	Human adeno-associ
23	50	10.9	7011	1	V20464	Human L-myc oncov
24	49.6	10.9	1303	1	Q23297	HSV-1 (CVG-2) ICP3
25	49.2	10.8	701	1	Q76209	HSV L/ST ORF1. Her
26	49.2	10.8	117213	1	V62176	HSV-2 strain SB5 C
27	49	10.7	4405	1	Q25444	PACE composite seq
28	49	10.7	4405	1	T05556	Human PACE coding
29	49	10.7	4488	1	T35520	Human thrombopoiet
30	49	10.7	223	1	T66412	Wild type SNF 5' u
31	49	10.7	223	1	V01700	Human T-cell leuka
32	48.4	10.6	801	1	V73801	KSHV LTR terminal
33	48	10.5	4020	1	T91361	OSF virus genomic
34	48	10.5	117213	1	V62176	HSV-2 strain SB5 C
35	47.8	10.5	1032	1	O44281	Pseudomonas cepac
36	47.6	10.4	2823	1	T35233	Natural killer cyt
37	47.6	10.4	12700	1	V62133	HSV-2 strain SB5 C
38	47.6	10.4	11705	1	V62130	HSV-2 strain SB5 C
39	47.4	10.4	1610	1	O57657	Allele D4.7 of the
40	47.4	10.4	1608	1	T27547	Recombinant human
41	47.4	10.4	4425	1	V32641	Human chorion cdna
42	47.4	10.4	16812	1	V62175	HSV-2 strain SB5 C
43	47.2	10.3	4356	1	O37543	Cardiac adenylyl c

# ALIGNMENTS

44 47.2 10.3 4356 1 095540  
45 47 10.3 8438 1 073500

ASIS:

RESULT 1	T48239	standard; DNA; 1876 BP.
ID	T48239	
AC	T48239	
DT	09-APR-1997	(first entry)
DE	Neuronal nicotinic acetylcholine receptor alpha	
KW	Neuronal nicotinic acetylcholine receptor; nAChR; neurotransmitter;	
OS	ligand-gated receptor; ds.	
OS	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	cds	73..1581
FT	cds	/*tag= a
PN	W09641876-A1.	
PD	27-DEC-1996.	
PF	07-JUN-1996; U09775.	
PR	07-JUN-1995; US-484722.	
PA	(SIBI-) SIBIA NEUROSCIENCES INC.	
PI	Elliot KJ, Harpold KM;	
DR	WPI; 97-065463/06.	
P-PSDB	W09025.	
PT	Nucleic acids encoding neuronal acetylcholine receptor sub-units -	
PT	used in screening to determine the effect of drugs on the receptor	
PS	disclosure; Page 71-73; 108pp; English.	
CC	A DNA sequence (T48239) codes for the alpha-7 subunit (W09025) of	
CC	the human neuronal nicotinic acetylcholine receptor (nAChR). Host	
CC	cells, esp. mammalian cells or amphibian oocytes, carrying alpha-7	
CC	nucleic acids, opt. in combination with other alpha and/or beta	
CC	subunit nucleic acids (see also T48232-38, T48240-41), express	
CC	recombinant nAChR subunits useful for identifying cpds. that	
CC	modulate the activity of human nAChRs.	
SQ	Sequence 1876 BP; 369 A; 553 C; 531 G; 423 T;	
Query Match 27.8%; Score 127; DB 1; Length 1876;		
Best Local Similarity 100.0%; Pred. No. 2.2e-13;		
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY	321 GGGCGGAGGCGGCGGCGGCGGAGCGGAGCGGCGGCGGCGGCGGCTCGTCACT	380
DB	1 GGGCGGAGGCGGCGGCGGCGGCGGAGCGGAGCGGCGGCGGCGGCTCGTCACT	60
OY	381 CCGGAGCTCAACATGCGCTGCTGCGGAGGCGGCTGCTGCGGCGGCGGCTC	440
DB	61 CCGGAGCTCAACATGCGCTGCTGCGGAGGCGGCTGCTGCGGCGGCGGCTC	120
OY	441 CTGCACG 447	
DB	121 CTGCACG 127	
RESULT 2		
ID	V12197	standard; cDNA; 1876 BP.
AC	V12197	
DT	14-MAY-1998	(first entry)
DE	Human neuronal nicotinic acetylcholine receptor alpha-7 subunit cDNA.	
KW	Human; neuronal nicotinic acetylcholine receptor; alpha-7 subunit;	
KW	brain tissue; screening; NACR; antibody; ds.	
OS	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	cds	73..1581
FT	cds	/*tag= a
FT	cds	/product= "neuronal nicotinic acetylcholine receptor
PN	W09420617-A2.	
PD	15-SEP-1994.	
PD	alpha-7 subunit"	

PF 08-MAR-1994: U02447.  
 PR 08-MAR-1993; US-028031.  
 PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.  
 PA (SIBI-) SIBIA NEUROSCIENCES INC.  
 PI Elliott KJ, Ellis SB, Harpold MM;  
 DR WPI: 94-303024/37.  
 DR P-PSDB: W44153.  
 PT Human neuronal nicotinic acetylcholine receptor subunits and DNA -  
 PT also transformed cells useful for screening cpds. which modulate  
 PT activity of the receptor  
 PS Claim 8; Page 78-79; 99pp; English.  
 CC The present sequence encodes a human neuronal nicotinic acetylcholine  
 CC receptor (NACHR) subunit. The cells expressing the alpha and/or beta  
 CC NACHR subunits may be used in a method of screening compounds to  
 CC identify any which modulate the activity of human neuronal NACHR.  
 CC Subunit specific antibodies may be used to monitor the distribution  
 CC and expression density of various subunits in normal vs diseased brain  
 CC tissues. Testing of single receptor subunits or specific receptor  
 CC subunit combinations with a variety of potential agonists or antagonists  
 CC provides information with respect to the function and activity of the  
 CC individual subunits and should lead to the identification and design of  
 CC compounds that are capable of very specific interaction with one or  
 CC more receptor subtypes. The resulting drugs should exhibit fewer  
 CC unwanted side effects than drugs identified e.g. screening with cells  
 CC that express a variety of subtypes.  
 CC Sequence 1876 BP; 369 A; 553 C; 530 G; 423 T;  
 SQ

Query Match 27.8%; Score 127; DB 1; Length 1876;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-13;  
 Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 321 GCGCCGAGCGGAGCGCCGCGGAGACCGAGACGTGAGCGCGCGCTCCTCAGCT 380  
 DB 1 GCGCCGAGCGGAGCGCCGCGGAGACCGAGACGTGAGCGCGCGCTCCTCAGCT 60  
 QY 381 CCGGAGCTCAATCGCTGCTCGCGGAGCGCTGCTGCTGCGCGCGCTGCTGCTC 440  
 DB 61 CCGGAGCTCAATCGCTGCTCGCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTC 120  
 QY 441 CTGCACG 447  
 DB 121 CTGCACG 127

RESULT 3  
 051543/c  
 051543 standard; DNA: 3861 BP.  
 051543:  
 24-MAY-1994 (first entry)  
 DE Lipopolysaccharide induced protein gene.  
 KW Macrophage; induced; lipo-polysaccharide; antitumor;  
 KW antiinflammatory; trypanocidal agent; antibody; cell proliferation;  
 KW activation; cytotoxicity; ds.  
 OS Homo sapiens.  
 FH Key  
 FT cds Location/Qualifiers  
 FT 1810..3315  
 FT /tag- a  
 FT 1980..2228  
 FT /tag- b  
 FT /note- "Intron 1 is only partially sequenced  
 FT (represented by (N)40 in the sequence  
 FT estimated length ca. 5400 bp"  
 FT 2615..2846  
 FT /tag- c  
 FT /note- "Intron 2 is only partially sequenced  
 FT (represented by (N)40 in the sequence  
 FT estimated length ca. 7900 bp"  
 FT 2907..2993  
 FT /tag- d  
 FT /note- "Intron 3 is only partially sequenced  
 FT (represented by (N)40 in the sequence  
 FT estimated length ca. 1000 bp"  
 FT Intron  
 FT Intron

FT poly-a-signal 3841..3846  
 FT /tag- e  
 FT W09322437-A.  
 PD 11-NOV-1993.  
 PR 28-APR-1993; E01022.  
 PR 30-APR-1992; E010231.  
 PA (INNO-) INNOGENETICS NV SA.  
 PI Devos K, Franssen L, Van De Voorde A, Van Heuverswyn H;  
 DR WPI: 93-368796/46.  
 DR P-PSDB: R43682  
 PT New polypeptide induced in macrophage(s) by lipo-polysaccharide -  
 PT useful e.g. as antitumor, antiinflammatory or trypanocidal  
 PT agent, also related nucleic acid, antibodies, anti-sense cpds.  
 PT etc.  
 PS Claim 9; Fig 1; 108pp; English.  
 CC The polypeptide induced in macrophages by lipopolysaccharide  
 CC stimulates cell proliferation (esp. when costimulated with IL-4)  
 CC promote activation, cytotoxicity, and mobilisation of LAK cells;  
 CC promote recruitment of suppressive peritoneal exudate cells;  
 CC promote generation of immunocompetent lymph node cells (LNC) and  
 CC have trypanocidal and trypanolytic activity. The human and murine  
 CC sequences are given in (051543-45), peptide fragments able to  
 CC generate antibodies are given in (R51951-61)  
 CC Sequence 3861 BP; 542 A; 1241 C; 1292 G; 658 T;  
 SQ

Query Match 15.1%; Score 69.2; DB 1; Length 3861;  
 Best Local Similarity 53.3%; Pred. No. 0.00033;  
 Matches 146; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 35 CAGCCCTGCTGCTGCGCAGAGCGCGAGCGCGAGAGCCCGCTCGTGAGACTGGGGTG 94  
 DB 2086 CTGCCCG 2027  
 QY 95 GAGTGCCTCGGAGCGTACACCGCGCGGAGATCTCCGCTACACCTCGGCTGCACT 154  
 DB 2026 GAGCGGAGGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1967  
 QY 155 TCCCTGTGATGCGCGCGAGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 214  
 DB 1966 TGCACCGGTGCTGGAATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1907  
 QY 215 GCGGGGCGGCGGCTGCTACGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 274  
 DB 1906 GCACGACGCGGAGCGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1847  
 QY 275 CCGCGCTCTTAAAGCGCGCGCGAGCGCGCG 308  
 DB 1846 CCG 1813

RESULT 4  
 X53491  
 ID X53491 standard; DNA: 114955 BP.  
 AC X53491;  
 DT 05-JUL-1999 (first entry)  
 DE Human adenovine A1 receptor antisense oligonucleotide fragment.  
 KW Antisense oligonucleotide; multiple target; antisense treatment;  
 KW impaired respiration; inflammation; lung disease;  
 KW pulmonary vasoconstriction; inflammation; allergic rhinitis;  
 KW acute asthma; allergy; asthma; impeded respiration;  
 KW respiratory distress syndrome; pain; cystic fibrosis;  
 KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;  
 KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma.  
 KW colon cancer; breast cancer; lung cancer; pancreatic cancer;  
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metast  
 KW prostate cancer; ss.  
 OS Synthetic.  
 PN W09913886-A1.  
 PD 25-MAR-1999.  
 PR 17-SEP-1998; U19419.  
 PR 09-JUN-1998; US-083972.  
 PR 17-SEP-1997; US-059160.

(UYEC-) UNTV EAST CAROLINA.

Nyce JW; 99-229400/19.  
PI New antisense oligonucleotides used in treatment of, e.g., pulmonary  
PT vasoconstriction

Pt Disclosure; Page 37; 120pp; English.

PS The specification describes antisense oligonucleotides (X52866-X55271)  
CC directed against at least 2 mRNAs selected from target genes, coding and  
CC non-coding regions of RNAs corresponding to target genes, gene  
CC initiation codons, genomic flanking regions, intron-exon borders, the  
CC 5'-end, the 3'-end and the juxta-section between coding and non-coding  
CC regions and all segments of RNAs encoding proteins associated with one  
CC or more diseases, conditions or mixtures. The antisense oligonucleotides  
CC may be derived from sequences X55272-74. These multiple target  
CC oligonucleotides (specifically X55180-271) can be used for the antisense  
CC treatment of diseases and conditions. Typical diseases and conditions  
CC are those associated with impaired respiration and inflammation,  
CC including lung diseases, pulmonary vasoconstriction, inflammation,  
CC allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,  
CC respiratory distress syndrome, pain, cystic fibrosis, pulmonary  
CC hyperinflation, pulmonary vasoconstriction, emphysema, chronic obstructive  
CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,  
CC carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic  
CC cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic  
CC metastases, as well as all types of cancers which may metastasize or have  
CC metastasized to the lungs, including breast and prostate cancer.

SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T;

Query Match            15.1%; Score 68.8; DB 1; Length 114955;  
Best Local Similarity .37.3%, Pred. No. 0.0002;  
Matches 156; Conservative 33; Mismatches 227; Indels 2; Gaps 1

QY 33 GGCAACCCCCCTGGGCTCGAGGCAGAGGCCGAGGCGCGAAGACCCTGTGATGAAGTCGGGG 92  
||| ||| ||| ||| | : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 108466 GCGCGCGCTTGCTCTCSNNNDNNGGCGGGGGCGCGCGTCTGACTTSNNNDNNGCGCGGG 108525

QY 93 TGAGAGTTCCCGGAGCGTAGC--CACGCGCGGAGAATTACTCCCCTCATACACTCGGGCTG 150  
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 108526 CGCGCGCGCTTGCTGGSNNNDNNGCGGGGGCGCGCGGCGCTGGSNNNDNNGCGGGGGCG 108585

QY 151 CAATTCCCTTGGTGTGGCCCCCGAGAGAGCTGTGCCCGGGCTGGAAGGATGCGGGGCGGGAGC 210  
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 108586 CGCGCGCTGSSNNNDNNGCGGGGGCGCGCGCGCTTSNNNDNNGGCGGGGCGCGCGCC 108645

QY 211 GGGGCGGGGGCGGGGCGCTGTCAGCTGTGAGAGCGCGCGGGGGCGGGCGGGCGGGGGCG 270  
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
108646 SNNNDNNGCGGGGGCGGGCGCGCGSNNNDNNGCGGGGGCGGGCGSGSNNDNNGSGCGGG 108705

QY 271 CGCGCCCGGCTCTTTAAAGGCGCGGCGAGACCGAGCGAGAGTGTCTGTGTGAGCCAGAC 330  
:||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 108706 GCGCGCGCGSNNNDNNGCGGGGGCGGGGSGNSNNDNNGCGGGGGCGCGGSGSNNDNNGGC 108765

QY 331 GCAGAGCCCGGGCGAGACCGGAGAGCTGTGAGCGCGCGGCTGCTGTACGTCTCGGAGACTCA 390  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 108766 GGGGCGCGSNNNDNNGCGGGGGCGGSGNSNNDNNGCGGGGGCGCGCGCTGTCTGTGCGC 108825

QY 391 ACATGCGCTGTCTCGCGGAGAGCGTGTGCGGCTGTGCGCGCGTCTCTGTGAGCGG 448  
||| ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 108826 TBGGGCGCCSNNNDNNGCGGGGGCGCGCGCTGTGCTGTGCGGCGCCSNNNDNNG 108883

RESULT 5  
ID X53491 standard; DNA; 114955 BP:  
AC X53491;  
DT 05-JUL-1999 (first entry)

DE Human adenovirus AI receptor antisense oligonucleotide fragment.  
KW Antisense oligonucleotide; multiple target; antisense treatment;  
IM Impaired respiration; Inflammation; Lung disease;  
PM Pulmonary vasoconstriction; Inflammation; Allergic rhinitis;  
KM Acute asthma; allergy; asthma; impeded respiration;  
KW respiratory distress syndrome; pain; cystic fibrosis;

KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;  
 KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;  
 KW colon cancer; breast cancer; lung cancer; pancreatic cancer;  
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;  
 KW prostate cancer; ss.  
 OS Synthetic.  
 PN W09913886-A1.  
 PD 25-MAR-1999.  
 PF 17-SEP-1998; U19419.  
 PR 09-JUN-1998; US-093972.  
 PR 17-SEP-1997; US-059160.  
 PA (UYEC-) UNIV EAST CAROLINA.  
 PI NYCE JW;  
 DR WPI; 99-229400/19.  
 PT New antisense oligonucleotides used in treatment of, e.g. pulmonary  
 PT vasoconstriction  
 PS Disclosure; Page 37; 120pp; English.  
 CC The specification describes antisense oligonucleotides (X52869-X55271)  
 CC directed against at least 2 mRNAs selected from target genes, coding and  
 CC non-coding regions of RNAs corresponding to target genes, gene  
 CC initiation codons, genomic flanking regions, intron-exon borders, the  
 CC 5'-end, the 3'-end and the juxta-section between coding and non-coding  
 CC regions and all segments of RNAs encoding proteins associated with one  
 CC or more diseases, conditions or mixtures. The antisense oligonucleotides  
 CC may be derived from sequences X55272-74. These multiple target  
 CC oligonucleotides (specifically X55180-271) can be used for the antisense  
 CC treatment of diseases and conditions. Typical diseases and conditions  
 CC are those associated with impaired respiration and inflammation,  
 CC including lung diseases, pulmonary vasoconstriction, inflammation,  
 CC allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,  
 CC respiratory distress syndrome, pain, cystic fibrosis, pulmonary  
 CC hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive  
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,  
 CC carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic  
 CC cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic  
 CC metastases, as well as all types of cancers which may metastasize or have  
 CC metastasized to the lungs, including breast and prostate cancer.  
 CC Sequence 114935 BP; 6071 A; 29417 C; 36712 G; 21328 T;

Query Match	Best Local Similarity	14.3%;	Score 65.2;	DB 1	Length 114955;
Matches 149;	Conservative 49;	Mismatches 221;	Indels 6;	Gaps	
QY	22	GCTTGCCCTTGGGCGAGCCCTGGCTGGCCAGAGGCGCGAGGCCGAGAGCCCGCTGGTG	81		
Db	104995	GCCCCGCGCGGCGCGCGCCCGCCVCGNNNNNNNSCGGCGCCGCGCGGCGCCCGCCVGGC	104936		
QY	82	GAGACTGGGGGTGAGAGTGTGCCCGGAGCGGTACCCAGCGCCGGGAGTA-----CCTCCCGC	135		
Db	104935	CVGCGNNNNNSCGGCCCCGGCGGCGCGCGCCCGCCVGGCGVCGGNNNNNSGCGCCVGGGNN	104876		
QY	136	TCAACACTCGGCTGACGTATCCCTGGGTGGCCGCGAGAGCTGCGCCGGGCTGGAGGA	195		
Db	104875	HNNNSVGGCCVGGCGNNNNNSCIVGGCCVGGCGNNNNNSCCTVGGCCVGGCGNNNNNSC	104816		
QY	196	TGGCGGGGCGGGGACGCGGGGCGGGGCGCGGCGCTGTCACGTGAGAGCGCGCGGGGCGG	255		
Db	104815	CCVGGCCVGGCGNNNNNSGCCCGCCVGGCCVGGCGGNNNNNSCGCGCCVGGCGVGGGNNHN	104756		
QY	256	GCGCGGGGCGGGGCGCGGCCCGGCTCTCTAAAGGCGGCGCGAGGCCGAGGGGAGGTGCC	315		
Db	104755	NSCGCGCCCTVGGCCVGGCGGNNNNNSCGCGCCCGCCVGGCCVGGCGGNNNNNSGCGCGCCVGG	104696		
QY	316	TCTGTGCGCGCAGGCGCAGAGCCCGGCGAGCGCGAGAGCTGAGAGCGCGCGCTCTGTG	375		
Db	104695	CCVGGCGNNNNNSGCGCGCCCGCCVGGCGCIVGGGNNNNNSCGGCGCGGCCVGGCGG	104636		
QY	376	CACCTCGGGGACACACATGCGTGTCTCCCGGAGAGGCTCTTGGCTGAGCGCTGGCGCGCT	435		
Db	104635	NNNNNSGCGGCGCGCCCGCCVGGCCVGGCGGNNNNNSGCGCGCGCCCGCCVGGCCVGGCGGNN	104576		
QY	436	CGCTC	440		







FT misc\_rna 370..372  
 FT /\*tag= d  
 FT /note= "first codon of ORF-1"  
 PN WO9428156-A.  
 PD 08-DEC-1994.  
 PF 20-MAY-1994; U05770.  
 PR 20-MAY-1994; US-065146.  
 PA (DAND ) DANA FARBER CANCER INST INC.  
 PI Schaffer PA, Yeh L.  
 DR WPI: 95-022825/03.  
 PT Herpes Simplex Virus (HSV) specific junction spanning transcript  
 PT - for inhibiting HSV L/ST synthesis, in the treatment of HSV  
 PT infection.  
 PS Disclosure: Page 38-44; 64pp; English.  
 CC An HSV-specific junction-spanning transcript (L/ST) maps at the 5'  
 CC end to the b repeat sequences of HSV DNA at approx. 3 and 125 kb,  
 CC and extends into the c repeat sequences of HSV DNA. Compounds  
 CC that inhibit HSV L/ST synthesis may be used as anti-HSV  
 CC virucides. The HSV-1 DNA sequence in the region of the L/STs  
 CC is given in Q76213.  
 Sequence 12001 BP: 1568 A; 4557 C; 4315 G; 1561 T;  
 Query Match 12.3%; Score 56.4; DB 1; Length 12001;  
 Best Local Similarity 50.3%; Pred. No. 0.03; Mismatches 161; Indels 1; Gaps 1;  
 Matches 164; Conservative 0; Mismatches 161; Indels 1; Gaps 1;  
 QY 43 GGCCGCGCCAGAGCGCGGAGAGCCCGCTGCTGTGAGACTGGGGTGTGAGTGC 102  
 Db 2319 GGGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGG 2260  
 QY 103 CGGACGCTACCCAGCGCGGAGTACTCCGCTCACACTCTGGGCTGCACTTCC 162  
 Db 2259 CGAGGGGGCGGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGG 2200  
 QY 163 TGGCCCGCGAGACGCTGCGCGCTGAGAGGATGCGGGGGGAGAGGGGGCGG 222  
 Db 2199 GAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGG 2140  
 QY 223 GGGGCTGCTGACGTGAGAGAGCGCGGGGGCGGGGGCGGGGGCGCGCCGCTG 282  
 Db 2139 GAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGG 2080  
 QY 283 CTTAAGGCGCGGAGCGGAGCGGAGGTGCTCTGTGGCCGAGCGGAGCGCGG 342  
 Db 2079 GAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGG 2021  
 QY 343 GACAGCCGAGACGTGAGCGCGCGG 368  
 Db 2020 GAGGGGGCGAGGGGGCGAGGGGGG 1995  
 RESULT 10  
 ID V60339 standard; cDNA; 1120 BP.  
 AC V60339;  
 DT 04-DEC-1998 (first entry)  
 DE cDNA sequence of fibroblast growth factor-2 (FGF-2).  
 KW Fibroblast growth factor-2; FGF-2; leaderless protein; inhibition;  
 KW export; angiogenesis; restenosis; treatment; tumour; inflammation;  
 KW cell proliferation; diabetes; retinopathy; infection;  
 KW polycystic kidney disease; atherosclerosis; ss.  
 OS Homo sapiens.  
 PN WO9837880-A1.  
 PD 03-SEP-1998.  
 PF 25-FEB-1998; U03689.  
 PR 26-FEB-1997; US-807014.  
 PA (CIBL-) CIBLEX CORP.  
 PI Baird A, Florkiewicz RZ;  
 DR WPI: 98-495377/42.  
 PT Inhibiting export of leaderless protein with agent that inhibits  
 PT binding to transporter protein - especially for treating  
 PT angiogenesis and restenosis by preventing export of fibroblast

PT growth factor, also methods for identifying leaderless proteins and  
 PT their transporters  
 PS Claim 2; Pages 53-54; 116pp; English.  
 CC The present sequence encodes fibroblast growth factor-2 (FGF-2), a  
 CC leaderless protein. A leaderless protein refers to a protein that is  
 CC found in an extracellular environment, but lacks a canonical leader  
 CC sequence. The specification describes a method for inhibiting export of  
 CC a leaderless protein from a cell. The method comprises treating  
 CC the cell with an agent that inhibits binding between the leaderless  
 CC protein and a transport molecule. Treatment with the inhibiting agent  
 CC is specifically used to treat angiogenesis and restenosis, i.e. where  
 CC expression of FGF-2 is inhibited, and the agent is applied to endothelial  
 CC or smooth muscle cells. Other applications are treatment of tumours  
 CC (melanoma, teratocarcinoma, ovarian carcinoma, bladder cancer and  
 CC neuroblastoma), inflammation, cell proliferation, complications of  
 CC diabetes (e.g. retinopathy), viral, bacterial or fungal infections,  
 CC polycystic kidney disease and atherosclerosis.  
 SQ Sequence 1120 BP: 260 A; 293 C; 342 G; 225 T;  
 Query Match 12.3%; Score 56; DB 1; Length 1120;  
 Best Local Similarity 47.7%; Pred. No. 0.055;  
 Matches 164; Conservative 0; Mismatches 180; Indels 0; Gaps 0;  
 QY 41 CTGGCTGGCCAGAGGGCGGAGAGCCCGCTGCTGGTGGAGACTGGGGTGGAGTG 100  
 Db 83 CGGCTGCCAGAGGGCGGCTGCCGGAGACACCATCGTGAACCCAGGTCGGGGCG 142  
 QY 101 CCCGAGACGTACCCAGCGCGGAGTACTCCGCTCACACTCGGGCTGCACTTCC 160  
 Db 143 CCGGCTGCGCGCGACCCAGGGGGCGGGCGGAGACAGAGCGCGGCTGAGGCTG 202  
 QY 161 GTGGCGCGGAGAGCTGCGCGCTGAGAGGATGCGGGGGCGGAGCGGGGGG 220  
 Db 203 GGGAGCGGGGGCGGCGCGCGCTGCGGGCGGAGAGGCTGCGGGGGCGGGGGG 262  
 QY 221 GCGGGGCTGCTGACGTGAGAGAGCGCGCGGGGGCGGGGGCGGGGGCGCGCG 280  
 Db 263 CCGTGGCCCGAGGGGTGAGAGCGCGGGGGCGGGGGCGGGGGCGGCTCCCGG 322  
 QY 281 TCCTTAAGGCGCGGAGCGGAGCGGAGGTGCTCTGTGGCCGAGCGGAGCGCGG 340  
 Db 323 CGGCTCAGCGGCTCGGGGATCCGGCGGGGGCGGCGGAGGACATGGACCGCGG 382  
 QY 341 GCGAGCCGAGAGAGCTGAGAGCGCGCGCTGCTGAGACTCCG 384  
 Db 383 TCACACAGCTCCCGCTTGGCCCGGAGGATGGCGGAGCGGCGG 426  
 RESULT 11  
 ID V10362/c  
 AC V10362;  
 DT 30-JUN-1998 (first entry)  
 DE Infected cell protein number 4 alpha-4 gene.  
 KW Infected cell protein number 4; ICP4; alpha-4; cell apoptosis;  
 KW therapeutic; ss.  
 OS Herpes simplex virus.  
 FH Key Location/Qualifiers  
 FT CDS 361..4257  
 FT /\*tag= a  
 FT /product= "Infected cell protein"  
 PN WO9804709-A2.  
 PD 05-FEB-1998.  
 PF 23-JUL-1997; U12904.  
 PR 26-JUL-1996; US-690473.  
 PA (ARCH-) ARCH DEV CORP.  
 PI Leopardi R, Roizman B;  
 DR WPI: 98-130697/12.  
 DR P-PSDB; W40200.  
 PT Use of herpes simplex virus ICP4 polypeptide - useful for, e.g.  
 PT blocking apoptosis in cells, production of proteins and gene therapy  
 PS Disclosure; Fig 2; 63pp; English.

CC The herpes simplex virus (HSV) alpha-4 gene encodes infected cell protein  
 CC number 4 (ICP4). Both the alpha-4 gene and ICP4 can be used to block  
 CC cell apoptosis. Similarly the administration of an agent that inhibits  
 CC ICP4 or the alpha-4 gene can induce apoptosis in HSV infected cells.  
 CC This can be used for the immortalisation of cells, production of  
 CC proteins, gene therapy, or inhibition of cell death induced in vivo.  
 CC They can also be used for production of therapeutics comprising  
 CC inhibitors of HSV ICP4 function, useful for treating HSV infection.  
 SQ Sequence 4257 BP; 412 A; 1768 C; 1663 G; 414 T;

Query Match 11.6%; Score 53; DB 1; Length 4257;  
 Best Local Similarity 48.7%; Pred. No. 0.13;  
 Matches 172; Conservative 0; Mismatches 180; Indels 1; Gaps 1;

QY 88 GGGGCTGAGAGTCCCGGAGCCGAGACCGGAGTACCTCCGCTACACCTCGGG 147  
 DB 2751 GGGGCTGAGAGTCCCGGAGCCGAGACCGGAGTACCTCCGCTACACCTCGGG 147  
 148 CTGCACTTCCTGAGGCGGCGGAGACGCTGCGGCGGAGGAGTGGCGGCGG 207  
 DB 2691 CTGCGGCGGCGGCGGAGGCGGAGGCGGCGGAGGAGGAGGAGGCGG 207  
 QY 208 GACGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGAGGAGGAGGCGG 267  
 DB 2631 GCGGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGAGGAGGAGGCGG 267  
 QY 268 GCGGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGAGGAGGAGGCGG 2572  
 DB 2571 GCGGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGAGGAGGAGGCGG 2572  
 QY 328 GCGGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGAGGAGGAGGCGG 387  
 DB 2511 GCGGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGAGGAGGAGGCGG 387  
 QY 388 TCACATGCGCTGCTGCGGAGGCGGAGGCGGAGGCGGAGGAGGCGG 440  
 DB 2452 GCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2400

RESULT 12

V68520/c  
 ID V68520 standard; DNA; 4257 BP.  
 AC V68520;

DE 29-JAN-1999 (first entry)  
 The nucleotide sequence of the Herpes simplex virus ICP4.  
 Herpes simplex virus ICP4; HSV ICP4; modulator; apoptosis;  
 stimulation; inhibition; HSV infection; ss.  
 FH Herpes simplex virus

US Key Location/Qualifiers  
 CDS 361..4257

FT /tag= a  
 FT /product= "HSV ICP4 protein"

PN MO946637-A2.  
 PD 22-OCT-1998.  
 PE 16-APR-1998; U05753.  
 PR 16-APR-1997; US-843659.  
 PA (ARCH-) ARCH DEV CORP.  
 PI Leopardi R, Roizman B;  
 DR WPI; 98-594559/50.  
 DR P-PSDB; W80810.  
 PT Use of herpes simplex virus U(s)3 polypeptide - for developing  
 PT products for modulating apoptosis in cells and for identifying  
 PT compounds which act as stimulators or inhibitors of apoptosis  
 PS Example 2; Pages 60-63; 85pp; English.  
 CC This is the nucleotide sequence of Herpes simplex virus ICP4 used in  
 CC and products can be used to identify compounds of apoptosis. The methods  
 CC (stimulate or inhibit) apoptosis in cells. They can be used to  
 CC immortalise cells for the study of these cells or for growing cells  
 CC in large numbers for the production of proteins. They can also be  
 CC used for stimulating apoptosis in cells, e.g. for treating a subject  
 CC with a HSV infection.

SQ Sequence 4257 BP; 412 A; 1768 C; 1663 G; 414 T;

Query Match 11.6%; Score 53; DB 1; Length 4257;  
 Best Local Similarity 48.7%; Pred. No. 0.13;  
 Matches 172; Conservative 0; Mismatches 180; Indels 1; Gaps 1;

QY 88 GGGGCTGAGAGTCCCGGAGCCGAGACCGGAGTACCTCCGCTACACCTCGGG 147  
 DB 2751 GGGGCTGAGAGTCCCGGAGCCGAGACCGGAGTACCTCCGCTACACCTCGGG 147  
 148 CTGCACTTCCTGAGGCGGCGGAGACGCTGCGGCGGAGGAGTGGCGGCGG 207  
 QY 2691 CTGCGGCGGCGGCGGAGGCGGAGGCGGCGGAGGAGGAGGAGGCGG 207  
 DB 2631 GCGGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGAGGAGGAGGCGG 2632  
 QY 208 GACGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGAGGAGGAGGCGG 267  
 DB 2631 GCGGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGAGGAGGAGGCGG 267  
 QY 268 GCGGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGAGGAGGAGGCGG 2572  
 DB 2571 GCGGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGAGGAGGAGGCGG 2572  
 QY 328 GCGGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGAGGAGGAGGCGG 327  
 DB 2571 GCGGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGAGGAGGAGGCGG 2512  
 QY 2511 GCGGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGAGGAGGAGGCGG 387  
 DB 2452 GCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2453  
 QY 388 TCACATGCGCTGCTGCGGAGGCGGAGGCGGAGGCGGAGGAGGCGG 440  
 DB 2452 GCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2400

RESULT 13

Q76213  
 ID Q76213 standard; DNA; 12001 BP.  
 AC Q76213;  
 DT 04-AUG-1995 (first entry)  
 DE HSV L/ST region.

KW HSV; junction-spanning transcript; L/ST; therapy; virucide; ds.  
 OS Herpes simplex virus type 1.  
 FH Key Location/Qualifiers  
 FT 199..204  
 FT /tag= a  
 FT /tag= b  
 FT /note= "ICP4 binding site"

FT misc\_binding  
 FT 220..225  
 FT /tag= b  
 FT /note= "ICP4 binding site"

FT misc\_rna  
 FT 229  
 FT /tag= c  
 FT /note= "5' end of the L/STs"

FT misc\_rna  
 FT 370..372  
 FT /tag= d  
 FT /note= "first codon of ORF-1"

PN MO9428156-A.  
 PD 08-DEC-1994.  
 PE 20-MAY-1994; U05770.  
 PR 20-MAY-1993; US-065146.  
 PA (DAND) DANA FARBER CANCER INST INC.  
 PI Schaffer PA, Yeh L;  
 DR WPI; 95-022825/03.  
 DR Herpes Simplex Virus (HSV) specific junction spanning transcript  
 PT - for inhibiting HSV L/ST synthesis, in the treatment of HSV  
 PT infection.  
 PS Disclosure; Page 38-44; 64pp; English.  
 CC An HSV-specific junction-spanning transcript (L/ST) maps at the 5'  
 CC end to the b repeat sequences of HSV DNA at approx. 3 and 125 kb,  
 CC and extends into the c repeat sequences of HSV DNA. Compounds  
 CC that inhibit HSV L/ST synthesis may be used as anti-HSV  
 CC virucides. The HSV-1 DNA sequence in the region of the L/STs  
 CC is given in Q76213.  
 SQ Sequence 12001 BP; 1568 A; 4557 C; 4315 G; 1561 T;

Query Match

11.5%; Score 52.4; DB 1; Length 12001;



Sun Sep 19 10:27:05 1999

us-08-956-518a-94.rng

Page 9

Search completed: September 18, 1999, 00:34:06  
Job time: 19006 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:07 ; Search time 2825.05 seconds  
(without alignments)  
319.091 Million cell updates/sec

Title: US-08-956-518a-94  
Perfect score: 457  
Sequence: 1 AGAAGCAGAGGAGAGGTAG.....CTCCGACAGGTAAGCCAC 457

Scoring table: IDENTITY\_NUC  
Searched: 2546578 seqs, 986266752 residues

Database :	
1:	em_est1:*
2:	em_est2:*
3:	em_est3:*
4:	em_est4:*
5:	em_est5:*
6:	em_est6:*
7:	em_est7:*
8:	em_est8:*
9:	em_est9:*
10:	em_est10:*
11:	em_est11:*
12:	em_est12:*
13:	em_est13:*
14:	em_est14:*
15:	em_est15:*
16:	em_est16:*
17:	em_est17:*
18:	em_est18:*
19:	em_est19:*
20:	gb_est1:*
21:	gb_est2:*
22:	gb_est3:*
23:	gb_est4:*
24:	gb_est5:*
25:	gb_est6:*
26:	gb_est7:*
27:	gb_est8:*
28:	gb_est9:*
29:	gb_est10:*
30:	gb_est11:*
31:	gb_est12:*
32:	gb_est13:*
33:	gb_est14:*
34:	gb_est15:*
35:	gb_est16:*
36:	gb_est17:*
37:	gb_est18:*
38:	gb_est19:*
39:	gb_est20:*
40:	gb_est21:*
41:	gb_est22:*
42:	gb_est23:*
43:	gb_est24:*
44:	gb_est25:*
45:	gb_est26:*
46:	gb_est27:*
47:	gb_est28:*
48:	gb_est29:*
49:	gb_est30:*
50:	gb_est31:*
51:	gb_est32:*
52:	em_est20:*
53:	em_est21:*

54: em\_est22:\*\n
55: em\_est23:\*\n
56: em\_est24:\*\n
57: em\_est25:\*\n
58: em\_est26:\*\n

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	77.4	16.9	641	45	A1357868	A1357868 qv13b02.x
C 2	74.2	16.2	430	45	A1367590	A1367590 qg30a12.x
C 3	73	16.0	759	26	W28277	W28277 4498 Human
C 4	68.2	14.9	459	44	A1285778	A1285778 qu56b02.x
C 5	67.8	14.8	1461	25	W15723	W15723 mb53c01.r1
C 6	65.8	14.4	457	47	A1469170	A1469170 ta19d04.x
C 7	65	14.2	547	47	A1525191	A1525191 promrna-9
C 8	63	13.8	635	47	A1525004	A1525004 promrna-4
C 9	61.6	13.5	1208	26	W29297	W29297 mb99e03.r1
C 10	61.4	13.4	1075	26	W29419	W29419 mb98h11.r1
C 11	61	13.3	1171	26	W29436	W29436 mb99g07.r1
C 12	60.4	13.2	453	27	AA008005	AA008005 mg64c12.x
C 13	60.2	13.2	1161	26	W14959	W14959 mc68b10.r1
C 14	60.2	13.2	403	47	A1468375	A1468375 tg58d01.x
C 15	59.8	13.1	859	26	W28841	W28841 52d7 Human
C 16	59.6	13.0	1681	49	A1653819	A1653819 ty01h01.x
C 17	59.4	13.0	1243	26	W42204	W42204 mc69d11.r1
C 18	59.2	13.0	1242	26	W29459	W29459 mc03f02.r1
C 19	59.2	13.0	1145	27	AA028476	AA028476 m12le10.r
C 20	59.2	13.0	362	42	A1150327	A1150327 qf38a01.x
C 21	59.2	13.0	407	44	A1250825	A1250825 gx04f05.x
C 22	59	12.9	1311	25	W15735	W15735 mb53g03.r1
C 23	59	12.9	1084	27	AA028875	AA028875 mh90d12.x
C 24	59	12.9	1257	27	W98479	W98479 mg20c12.r1
C 25	58.8	12.9	1138	45	A1374006	A1374006 SMOVAFCAP
C 26	58.6	12.8	1162	26	W42215	W42215 mc69h10.r1
C 27	58.6	12.8	454	50	A1700029	A1700029 tt36d12.x
C 28	58	12.7	1291	25	W15737	W15737 mb53h03.r1
C 29	58	12.7	1111	26	W41963	W41963 mc68d10.r1
C 30	58	12.7	1056	26	W42205	W42205 mc69e09.r1
C 31	58	12.7	1087	27	AA023473	AA023473 mh74e06.x
C 32	58	12.7	454	45	A1340211	A1340211 q079f12.x
C 33	58	12.7	292	47	A1500717	A1500717 tn94b08.x
C 34	57.8	12.6	1136	25	W15902	W15902 mb56e10.r1
C 35	57.8	12.6	832	26	W28836	W28836 52d2 Human
C 36	57.8	12.6	418	48	A1611745	A1611745 tu77h12.x
C 37	57.4	12.6	1049	26	W42199	W42199 mc69c09.r1
C 38	57.2	12.5	743	26	W25793	W25793 13e7 Human
C 39	57	12.5	1222	26	W42209	W42209 mc69f11.r1
C 40	56.8	12.4	415	45	A1363984	A1363984 qv35b02.x
C 41	56.8	12.4	1104	26	W41968	W41968 mc68f11.r1
C 42	56.6	12.4	1286	26	W42216	W42216 mc69h11.r1
C 43	56.6	12.4	435	44	A1302382	A1302382 gpa49h12.x
C 44	56.2	12.3	1088	25	W15897	W15897 mb56b09.r1
C 45	56.2	12.3	829	26	W28819	W28819 52b8 Human

ALIGNMENTS

RESULT 1  
LOCUS A1357868 641 bp mRNA  
DEFINITION qv13b02.x1 NCI-GAP K168 Homo sapiens cDNA clone IMAGE:1981419 3'  
ACCESSION A1357868  
NTD g4109489  
VERSION A1357868.1 GI:4109489  
EST 15-FEB-1999  
Similar to TR:Q69340 Q69340 ORF1, ORF2, AND ORF3. ;, MRNA sequence.











Washington, Seattle, WA 98195  
Tel: 5106280100  
Fax: 5106280108  
Email: huanggm@yahoo.com.

## FEATURES

### Location/Qualifiers

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="bvtumor"
/notes="Organ: Prostate; Vector: pBluescript; Directional:
cDNA library was constructed using lambda Z' II kit
(Stratagene). mRNA was extracted from a fresh prostate
tumor tissues (Urology Department, University of
Washington)."

```

BASE COUNT	75 a	69 c	312 g	94 t	85 others
ORIGIN					

Query Match	13.88;	Score 63;	DB 47;	Length 635;
1st Local Similarity	43.98;	Pred. No. 0.21;		
atches 144; Conservative	0;	Mismatches 184;	Indels 0;	Gaps 0

QY	16	GGTGAGAGCTGCTGGCTTTGGGACAGCCCTGTGCTGTGCACAGAGCGCGGAGAGCCGC	75
Db	206	GGGAGNCANGAATAATTNGTGTGNCNCCGNNGTGGGANNNGNNGGANNAGAACCCGGGKTTTC	265
QY	76	TCGGTGGAGACTTGGGGGGTGGAGGTGTCCCGAGAGCTACCAGCGCCGGGAGTACCTCCCGC	135
Db	266	AGCAAGCGCNGCNGGGGGGNCANGAGNNGNCNCNGNNGNGGGCCNGGGGTTAAAGNGGGG	325
QY	136	TCACACTCGGGGCTGCAGATTCCCTGGGGTGCGCCGCGAGACGCTGGCCGGGATGGAGGA	195
Db	326	GGNNCGNCGGCGCNGGCNCGANGAGGGGTGGGNGNMGCGCNGTTCGGGGGACGGGGGGGGCGG	385
QY	196	TGGCGGGGCGGGGAGACGGGGGCGGGGGCGGGGCTGTCACTGTGAGAGAGCGCGCGGGGCG	255
Db	386	GGGGGGGGGGGGGNCGGGGGAGGGGGGGGGGGGGCGCCGTGGGGGGNGGCNCGCGGGGG	445
QY	256	GGCGGGGCGGGGCGCGCGCCGCGCTCTCTTAAAGCGCGCGAGCCGAGCCGAGCGGTGCC	315
Db	446	GGTGNGGNGNGGGGGCGGGGTGAGNGGGGGGCGCNCNAGNGAGAGGGGGGGGGGGGNG	505
QY	316	TCCTGTGCGCCGACGGCGCAGGCCCGGGCG	343
Db	506	TGGGNGGGGGGGGGGGGTGGTGGGGGAG	533

ULT	9
297	
LOCUS	
DEFINITION	
W29297	1208 bp mRNA EST 11-SEP-1996
mb996b3.r1	Soares mouse p3NMF19.5 Mus musculus cDNA clone
IMAGE:337564	5' similar to gb:J03750.rna2 Mouse single stranded
binding protein p9 mRNA, complete (MOUSE); mRNA sequence.	
mb2065	
accession	

TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Sep 1, 1995 this sequence version replaced

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:218964  
Seq primer: EMP1mer  
High quality sequence stop: 368.

## FEATURES

```

source
1. 1208
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:337554"
/clone_lib="Soares mouse p3MPr19.5"
/dev_stage="19.5 dpc total fetus"
/ldb_host="DH10b (ampicillin resistant)"
/note="Vector: pT713D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a GAG I - Oligo(dT) primer (5'
TGTTCACATCTGAAGTGGCGCGCGCTTTTCTTTTCTTTT 3'),
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."
BASE COUNT      253 a
ORIGIN           162 c      519 g      223 t      51 others

```

Query Match	13.5%	Score 61.6;	DB 26;	Length 1208;
Best Local Similarity	47.7%	Pred. No. 0.33;		
Matches 106; Conservative	0;	Mismatches 116;	Indels 0;	Gaps 0

Qy 147 GCTGCAGTTCCTGGGTGGCCCGCAGACGCTGGCCCGGCTGGAGGATGGCGGGGGCG 206  
||||| - - - - - ||| ||| ||| ||| |||  
Db 910 GCTGCATNNNNNNNNNGGGGNNNNNNNNNNNGGNGGGGNGGGGGGGGGGGGGG 969

[illegible][illegible]

REFERENCE  
AUTHORS  
1 (baes 1 to 1075)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.



QY 359 AGCGGCGCG 368  
 Db 1144 GCGCGCGGGG 1153

RESULT 12  
 LOCUS AA008005/c 453 bp mRNA EST 25-JUL-1996  
 DEFINITION mg64c12.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus CDNA  
 clone IMAGE:437782 5', mRNA sequence.

ACCESSION AA008005  
 NID 61463974  
 VERSION AA008005.1 GI:1463974  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 453)  
 REFERENCE 1 (bases 1 to 453)  
 JTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Gelsel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 The WashU-HMI Mouse EST Project  
 Unpublished (1996)  
 COMMENT On Nov 29, 1993 this sequence version replaced gi:430589.

TITLE  
 JOURNAL  
 COMMENT

CONTACT: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LINL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:263118  
 Seq primer: ETPPrimer  
 High quality sequence stop: 451.  
 Location/Qualifiers  
 1..453  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /map="22: 21"  
 /clone="IMAGE:437782"  
 /clone\_1lb="Soares mouse embryo NBME13.5 14.5"  
 /sex="unknown"  
 /tissue="embryo"  
 /dev\_stage="13.5-14.5dpc total fetus"  
 /lab\_host="DH10B"  
 /note="Vector: p773D-Pac (Pharmacia) with a modified  
 polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand CDNA  
 was primed with a Not I - oligo(dT) primer (5'  
 TGGTACCACTCTGAGTGGAGGCGCGCGGAGAAATTTTCTTTTCTTTTCTTTT  
 T 3'), on equal amounts of mRNA from 2 13.5dpc and 2  
 14.5dpc embryos (total RNA provided by Minori KO, Mayne  
 State Univ., from 2 ); double-stranded cDNA was ligated to  
 Eco RI adaptors (Pharmacia), digested with Not I and  
 cloned into the Not I and Eco RI sites of the modified  
 p773 vector. Library went through one round of  
 normalization, and was constructed by Bento Soares and  
 M.Fatima Bonaldo."

BASE COUNT 10 a 190 c 188 g 19 t 46 others  
 ORIGIN

Query Match 13.2%; Score 60.4; DB 27; Length 453;  
 Best Local Similarity 47.0%; Pired. No. 0.51;  
 Matches 171; Conservative 0; Mismatches 192; Indels 1; Gaps 1;

QY 22 GCGTGGCTTGGGAGCCCTGTGGCCAGAGCGCGAGGCCGAGACCCGCTCGTG 81  
 Db 444 GGCNCGCCGCGCGGCGGCGCCGCCGCGCNCGCCGCGGCGGCGCGCGCGCC 385

QY 82 GAGACTGGGGGTGAGGTGCCCGGAGGTACCAAGCCCGGAGTACTCTCCGTCACAC 141  
 Db 384 GCGCCCGCGCGGCGCGCGCGCGCGGCGCGGCGCGCGCGCGCGCGCGCGCG 325

QY 142 CTCGGGTGAGTCTCCCTGTGGTGGCGCGAGAGCGTGGCCCGCGGCGGATGGCGG 201  
 Db 324 CCGCNCNCG 265

QY 202 GCGGGGAGCGGGGCGGGGCGGGGCGGTCTGTACGTGAGAGCGCGCGCGCGCGCGG 261  
 Db 264 CGCCGNCGCCG 206

QY 262 GCGGGGCG 321  
 Db 205 CG 146

QY 322 GCGGAGCG 381  
 Db 145 GCG 86

QY 382 CGGG 385  
 Db 85 CGGG 82

RESULT 13  
 LOCUS W41959 1161 bp mRNA EST 21-MAY-1996  
 DEFINITION mc68b10.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus CDNA  
 clone IMAGE:353659 5', mRNA sequence.

ACCESSION W41959  
 NID 91326461  
 VERSION W41959.1 GI:1326461  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 1161)  
 REFERENCE 1 (bases 1 to 1161)  
 JTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Gelsel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 The WashU-HMI Mouse EST Project  
 Unpublished (1996)  
 COMMENT On Apr 14, 1993 this sequence version replaced gi:716741.

TITLE  
 JOURNAL  
 COMMENT

CONTACT: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LINL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:225459  
 Seq primer: ETPPrimer  
 High quality sequence stop: 370.  
 Location/Qualifiers  
 1..1161  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /map="7: 883A09: 13: 13q12.2-13q12.3"  
 /clone="IMAGE:353659"  
 /clone\_1lb="Soares mouse embryo NBME13.5 14.5"  
 /sex="unknown"

```

/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/Note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer 15',
TGTTCACATCTGAGAGTGAGAGCCGCGGAAATTTTCTTTTCTTTTCTTTT
T 3', on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M. Fatima Bonaldo."
BASE COUNT      185 a      251 c      470 g      208 t      47 others
ORIGIN

```

```

Query Match      13.2%; Score 60.2; DB 26; Length 1161;
Best Local Similarity 44.3%; Pred. No. 0.52;
Matches 140; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 26 GGCCTTGGGACGCCCCCTGCGCTGCCAGAGCCGCGAGACCCGCTCGGTGAGA 85
DB 843 GCCCTAGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
QY 86 CTGGGGGTGGAGGTGCGCCGAGCGCTACCCAGCGCGGGAGTACCTCCCGCTCACACTCG 145
DB 903 CGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 962
QY 146 GCGTCGAGTTCCTGGGTGCGCCGCGAGACGCTGCGCCGCGTGAAGATGGCGGGCG 205
DB 963 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1022
QY 206 GGGAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 265
DB 1023 GNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1082
QY 266 GGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 325
DB 1083 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1142
QY 326 CAGCGCAGAGCCCGGG 341
DB 1143 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1158

```

```

RESULT 14
AI468375 403 bp mRNA EST 30-MAR-1999
LOCUS t958d01.x1 NCI CGAP_Pt28 Homo sapiens cDNA clone IMAGE:2112961 3'
DEFINITION similar to contains element MskI repetitive element ;, mRNA
sequence.
ACCESSION AI468375
NID 94330465
VERSION AI468375.1 GI:4330465
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 403)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Mar 10, 1998 this sequence version replaced gi:294885.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.

```

```

Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LMW at:
www-bio.11ni.gov/bhrp/image/image.html

```

```

FEATURES
source
Insert Length: 784 Std Error: 0.00
Seq primer: -400p from Gdpc.
Location/Qualifiers
1..403
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:2112961"
/clone_id="NCI-CGAP_Pt28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"

```

```

Query Match      13.2%; Score 60.2; DB 47; Length 403;
Best Local Similarity 44.3%; Pred. No. 0.55;
Matches 116; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 185 GCGTGGAGGATGCGGGGGGGGAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 244
DB 402 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 343
QY 245 GCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 304
DB 342 GCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 283
QY 305 GCGAGGTGCTCTGTGGGCGGAGCGCGAGCCCGGCGACAGCCGAGAGCGCGG 364
DB 282 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 223
QY 365 CCGGCTGCTGAGCTCGGGAAGCTCAATGCGCTGCTGCCGGGAGGCGGTGGCTGCG 424
DB 222 CCCCCCCCCCCCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 163
QY 425 GCTGGCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 446
DB 162 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 141

```

```

RESULT 15
W28841/c 859 bp mRNA EST 08-MAY-1996
LOCUS W28841
DEFINITION 52d7 Human retina cDNA randomly primed sublibrary Homo sapiens
cDNA, mRNA sequence.
ACCESSION W28841
NID 91308789
VERSION W28841.1 GI:1308789
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 859)
AUTHORS Macke, J., Smallwood, P. and Nathans, J.
TITLES Adult Human Retina cDNA

```

## JOURNAL

Unpublished (1996)  
On May 9, 1995 this sequence version replaced g1:803988.

Contact: Dr. Jeremy Nathans  
Dr. Jeremy Nathans, Dept. of Molecular Biology and Genetics

Johns Hopkins School of Medicine

725 North Wolfe Street, Baltimore, MD 21205

Tel: 410 955 4678

Fax: 410 614 0827

Email: jeremy\_nathans@jhu.edu

Clones from this library are NOT available.

PCR Primers

FORWARD: CTTTGACCAATTCAGCCTGTTAGT

BACKWARD: GAGTGGCTTTATGATTTCTTCACAGGTAA

Seq primer: GGGTAAAGCAAGAAT.

Location/Qualifiers

1. 859

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="Human retina cDNA randomly primed sublibrary"

/sex="mixed (males and females)"

/tissue="retina"

/dev\_stage="adult"

/lab\_host="E. coli strain K802"

/note="Organ: eye; Vector: lambda gt10; Site\_1: EcoRI; Site\_2: EcoRI; The library used for sequencing was a sublibrary derived from a human retina cDNA library. Inserts from retina cDNA library DNA were isolated, randomly primed, PCR amplified, size-selected, and cloned into lambda gt10. Individual plaques were arrayed and used as templates for PCR amplification, and these PCR products were used for sequencing."

BASE COUNT 141 a 352 c 112 g 161 t 93 others

ORIGIN

## Query Match

Best Local Similarity 13.1%; Score 59.8; DB 26; Length 859;

Matches 103; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

```

QY 160 GGTGGCCCGCCGAGACGCTGCGCCGCGGCTGAGGATGCGCGCGGCGGACGCGCGCGG 219
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 823 GGGGNGGGGGGGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 764
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 220 GCGGGGCTCTGACGTGAGAGGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGG 279
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 763 GGGGGGGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 704
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

280 CTCCTTAAAGCGCGCGACGCGCGAGGTGCTCTGTGCGCGCGAGCGCGAGCGCGCGCG 339
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 703 GGGNNNGNGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 644
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 340 GCGCAGCGAGAGAGTGAAGCGCGCGCGG 368
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 643 GGGGGGGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 615
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: September 17, 1999, 21:28:10  
Job time: 14274 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:02:31; Search time 1962.17 Seconds

(Without alignments)  
(499.210 Million cell updates/sec)

Title: US-08-956-518a-95

Sequence: 1 CAGCGCGCCACATAGCTCC...CAACATTAGGTGATCGCC 308

Scoring table: IDENTITY\_NTC

Searched: 679419 seqs, 1590154680 residues

Database:

GenBank: \*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_cm.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pl1.\*  
8: gb\_pl2.\*  
9: gb\_pl3.\*  
10: gb\_pl4.\*  
11: gb\_pl5.\*  
12: gb\_pl6.\*  
13: gb\_pl7.\*  
14: gb\_pl8.\*  
15: gb\_pl9.\*  
16: gb\_pl10.\*  
17: gb\_pl11.\*  
18: gb\_pl12.\*  
19: gb\_pl13.\*  
20: gb\_pl14.\*  
21: gb\_pl15.\*  
22: gb\_pl16.\*  
23: gb\_pl17.\*  
24: gb\_pl18.\*  
25: gb\_pl19.\*  
26: gb\_pl20.\*  
27: gb\_pl21.\*  
28: gb\_pl22.\*  
29: gb\_pl23.\*  
30: gb\_pl24.\*  
31: gb\_pl25.\*  
32: gb\_pl26.\*  
33: gb\_pl27.\*  
34: gb\_pl28.\*  
35: gb\_pl29.\*  
36: gb\_pl30.\*  
37: gb\_pl31.\*  
38: gb\_pl32.\*  
39: gb\_pl33.\*  
40: gb\_pl34.\*  
41: gb\_pl35.\*  
42: gb\_pl36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	275.8	89.5	641	11	AF029838	Homo sapi

Result	Score	Match	Length	DB	ID	Description
2	275.8	89.5	689	11	AF029839	Homo sapi
3	205.2	66.6	291	9	HS2E10F	Homo sapi
4	181.2	58.8	298	9	HS2E10F	Homo sapi
5	181.2	58.8	137699	11	AC005630	Homo sapi
6	154.8	50.3	192439	35	AC007562	Homo sapi
7	146.4	47.5	178307	35	AC006328	Homo sapi
8	87.4	28.4	5281	7	HS008191	Homo sapi
9	51	16.6	1700	7	OSNR112	Homo sapi
10	50.4	16.4	171	9	HS165HF	Homo sapi
11	49.8	16.2	180	9	HS11532F	Homo sapi
12	49.4	16.0	189	9	HS11532F	Homo sapi
13	45	14.6	142123	11	AC004690	Homo sapi
14	40.8	13.2	2837	10	HS0G16P5	Homo sapi
15	39.6	12.9	8546	2	AF014839	Homo sapi
16	39.4	12.8	2991	8	AF036891	Homo sapi
17	39.2	12.7	182756	9	AC007789	Homo sapi
18	39	12.7	795	9	HS008191	Homo sapi
19	39	12.7	207466	35	AC007249	Homo sapi
20	38.8	12.6	2672	3	RABTHBP	Homo sapi
21	38.8	12.6	131649	11	AC005875	Homo sapi
22	38.4	12.5	3459	10	HS0A12582	Homo sapi
23	38.4	12.5	785	10	HS0A13727	Homo sapi
24	38.2	12.4	2992	5	A63308	Homo sapi
25	38.2	12.4	3289	10	HS0F1	Homo sapi
26	37.8	12.3	189356	11	AC005746	Homo sapi
27	37.6	12.2	152141	10	HS007000	Homo sapi
28	37.6	12.2	4510	12	HM000000	Homo sapi
29	37.4	12.1	9373	9	HM000000	Homo sapi
30	37.4	12.1	1111	9	HM000000	Homo sapi
31	37.4	12.1	8042	9	HM000000	Homo sapi
32	37.4	12.1	129327	10	HS171N11	Homo sapi
33	37.4	12.1	9043	10	HS000000	Homo sapi
34	37.4	12.1	2568	10	HS000000	Homo sapi
35	37.2	12.1	1919	3	BT081159	Homo sapi
36	37.2	12.1	2122	12	RN005989	Homo sapi
37	36.8	11.9	6945	2	AF047554	Homo sapi
38	36.8	11.9	42999	10	HS013369	Homo sapi
39	36.8	11.9	152261	17	HE1CG	Homo sapi
40	36.8	11.9	12979	17	HEHVS1SU	Homo sapi
41	36.8	11.9	26245	17	HS1US	Homo sapi
42	36.8	11.9	648	17	HS1US	Homo sapi
43	36.8	11.9	135360	35	AC006435	Homo sapi
44	36.8	11.9	177707	35	AC006450	Homo sapi
45	36.6	11.9	6183	1	STMORP	Homo sapi

## ALIGNMENTS

RESULT 1  
AF029838 641 bp mRNA 16-DEC-1998  
LOCUS Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE  
AUTHORS  
Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J.,  
Moore, T., Jacobs, S., Meriwether, J., Choi, M., Kim, E., Walton, K.,  
Bulling, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.  
Genomic organization and partial duplication of the human alpha7  
neuronal nicotinic acetylcholine receptor gene  
Genomics 52 (2), 173-185 (1998)  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,  
Breese, C., Davis, A., Hopkins, J. and Freedman, R.  
Direct Submission

JOURNAL Submitted (15-OCT-1997) Psychiatry, University of Colorado Health  
Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA

FEATURES  
Source

1. .641  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15q14"  
/clone="CHRNA7-DR1"  
/note="RACE product A/B/C/D"  
<1. .>641  
/note="alpha 7 neuronal nicotinic receptor mRNA; used in  
alternative splicing"

BASE COUNT 151 a 192 c 163 g 135 t

ORIGIN

Query Match 89.5%; Score 275.8; DB 11; Length 641;  
Best Local Similarity 98.7%; Pred. No. 2.1e-48;  
Matches 299; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 CAGCCGCCACATAGTCTCCGCCCAAGTCTCGGTGCTCCCTTGCATTTCACAGCCGCG-T 59  
DB 1 CAGCCGCCACATAGTCTCCGCCCAAGTCTCGGTGCTCCCTTGCATTTCACAGCCGCGT 60  
QY 60 CCCACGAGGCTACGCGGGGGGAGAGGTGAGAGCCCGAGAGTCCGGCGGGGGCCCCG 119  
DB 61 CCCACGAGGCTACGCGGGGGGAGAGGTGAGAGCCCGAGAGTCCGGCGGGGGCCCCG 120  
QY 120 CTTGTGCGCGCGCCATGACAGCGGCTCGGAGTCTCTTTCCGCGCCCTCCCGC 179  
DB 121 CTTGTGCGCGCGCCATGACAGCGGCTCGGAGTCTCTTTCCGCGCCCTCCCGC 180  
QY 180 CGAGGTGAGGAGAGATGTCATGTCAGAGTTCCAGGCCAACCGAAGTTACTGCGCT 239  
DB 181 CGAGGTGAGGAGAGATGTCATGTCAGAGTTCCAGGCCAACCGAAGTTACTGCG-CCT 239  
QY 240 CTATCTCCAGAGAACCCAGAGGACCAAGCCGCGCTCACGCCCCACCAATTAAAG 299  
DB 240 CTATCTCCAGAGAACCCAGAGGACCAAGCCGCGCTCACGCCCCACCAATTAAAG 299  
QY 300 TGA 302  
DB 300 TTA 302

LOCUS 2

AF029839 689 bp mRNA PRI 16-DEC-1998  
Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.

AF029839 93757794  
AF029839.1 GI:3757794

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Source

source

1. .689  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15q14"  
/clone="CHRNA7-DR2"  
/note="RACE product A/C/D"  
<1. .>689  
/note="alpha 7 neuronal nicotinic receptor mRNA; used in  
alternative splicing"

BASE COUNT 142 a 210 c 196 g 141 t

ORIGIN

Query Match 89.5%; Score 275.8; DB 11; Length 689;  
Best Local Similarity 98.7%; Pred. No. 2.1e-48;  
Matches 299; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 CAGCCGCCACATAGTCTCCGCCCAAGTCTCGGTGCTCCCTTGCATTTCACAGCCGCG-T 59  
DB 113 CAGCCGCCACATAGTCTCCGCCCAAGTCTCGGTGCTCCCTTGCATTTCACAGCCGCGT 172  
QY 60 CCCACGAGGCTACGCGGGGGGAGAGGTGAGAGCCCGAGAGTCCGGCGGGGGCCCCG 119  
DB 173 CCCACGAGGCTACGCGGGGGGAGAGGTGAGAGCCCGAGAGTCCGGCGGGGGCCCCG 232  
QY 120 CTTGTGCGCGCGCCATGACAGCGGCTCGGAGTCTCTTTCCGCGCCCTCCCGC 179  
DB 233 CTTGTGCGCGCGCCATGACAGCGGCTCGGAGTCTCTTTCCGCGCCCTCCCGC 292  
QY 180 CGAGGTGAGGAGAGATGTCATGTCAGAGTTCCAGGCCAACCGAAGTTACTGCGCT 239  
DB 293 CGAGGTGAGGAGAGATGTCATGTCAGAGTTCCAGGCCAACCGAAGTTACTG-CCT 351  
QY 240 CTATCTCCAGAGAACCCAGAGGACCAAGCCGCGCTCACGCCCCACCAATTAAAG 299  
DB 352 CTATCTCCAGAGAACCCAGAGGACCAAGCCGCGCTCACGCCCCACCAATTAAAG 411  
QY 300 TGA 302  
DB 412 TTA 414

RESULT 3

HSZELIOF

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Vector: pGEM-5zf(-)  
Clones are available from the UK MRC Human Genome Mapping Project  
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:  
http://www.hgmp.mrc.ac.uk/ for details  
or contact: biolhelp@hgmp.mrc.ac.uk.  
Location/Qualifiers  
1. .291



FEATURES	source
BASE COUNT	47 a 99 c 102 g 48 t 2 others
ORIGIN	
BASE COUNT	44 a 95 c 103 g 45 t 4 others
ORIGIN	
Query Match	66.6%; Score 205.2; DB 9; Length 291;
Best Local Similarity	96.9%; Pred. No. 9,6e-34;
Matches 218; Conservative	0; Mismatches 6; Indels 1; Gaps 1
QY	1 CAGCGCGCCACATAGCTCCCGCCAGATCCTCGATGCCCCCTTGCCATTTCAGCCGCG-T 59
DB	67 CAGCGCGCCACATAGCTCCCGCCAGATCCTCGATGCCCCCTTGCCATTTCAGCCGCGT 126
QV	60 CCCACGAGGGTACGCGCGCGGGGAGAGGTGTGAAGCCGACAGACTGCGCGGGGGCCCG 119
	127 CCCACGAGGGTACGCGCGCGGGGAGAGGTGTGAAGCCGACAGACTGCGCGGGGGCCCG 186
QY	120 CCTGATGCGCGCGCGCCATACAGCGGCTCGGGACATGCGCCCTTTCGCGGCCCTCCCG 179
DB	187 CCTGATGCGCGCGGNCATACAGCGGCTCGGGACAGGCTCTTTTCGCGGCCCTCCCG 246
QY	180 CGAGAGTGAGGGAGAGATGTCCATGTCAGAGGTTCAAGCCCAAC 224
DB	247 CGGAGTGAGGGAGAGATGTCCATGTCAGAGGTTCAAGCCCAAC 291
RESULT 4	
HS2G7F	298 bp DNA PRI 19-OCT-1995
LOCUS	H.sapiens Cpg Island DNA genomic MseI fragment, clone 2g7, forward
DEFINITION	H.sapiens Cpg Island DNA genomic MseI fragment, clone 2g7, fl.a.
ACCESSION	Z60646
NID	91032750
VERSION	Z60646.1 GI:1032750
KEYWORDS	Cpg Island; genomic MseI fragment.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS	Primates; Catarrhini; Homiidae; Homo.
JOURNAL	1 (bases 1 to 298)
TITLE	Macdonald, M., Huckle, E., Wilkinson, P. and Micklem, G.
REFERENCE	Direct Submission
AUTHORS	Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire
JOURNAL	CB10 1RQ, England. E-mail contact: humquerry@sanger.ac.uk
TITLE	2 (bases 1 to 298)
REFERENCE	Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.
AUTHORS	Purification of Cpg Islands using a methylated DNA binding column
JOURNAL	Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE	94282070
COMMENT	Vector: pGEM-5zf(-)
	Clones are available from the UK MRC Human Genome Mapping Project
	Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See url:
	http://www.hgmp.mrc.ac.uk/ for details
	or contact: bihelp@hgmp.mrc.ac.uk.
	Location/Qualifiers
	1. 298
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/sex="male"
	/dev_stage="adult"
	/tissue_type="blood"
	/clone_lib="CGI-1"
	/clone="2g7"

Query Match	58.8%	Score 181.2	DB 9	Length 288
Best Local Similarity	94.0%	Pred. No. 8.7e-29		
Matches 221	Conservative 0	Mismatches 9	Indels 5	Gaps 3
QY 4	CGCCGACATAGCTCCGCCCAAGTCTCGTGGCCCTTGCATTTTCCAGCCGG-TC	62		
DB 67	CGCCGACATAGCTCTCCGCCCAAGTCTCGTGGCCCTTGCATTTTCCAGCCGGCTCC	126		
QY 63	ACGAGGTCACGGCGCGGGGAGAGTGGAGCCGCAAGCTCGCCCGGGGCCCT	122		
DB 127	ACGAGGTCACGGCGCGGGGAGAGTGGAGCCGCAAMTCGGCC---GGGGCCCGCT	183		
QY 123	GTGGCCCCCGGCATGAGAGCGGCTCGGAGCTGCTTTTCGGCGGCCCTCCCGCCG	182		
DB 184	GGTGTCTCCGGCCAGACAGCGGCTCGGAGACAGCTCTTTTCGGCGGCCCTCCCGCCG	243		
QY 183	AGGTAGGGAAGATGTCCATGTCTAGGGTTCAGGCGCAACCGAAGTACTGGCC	237		
DB 244	A-GTGAAGGGAAGATGTCTCATGTCTCGGCTTCAAGGCCAACCGAAGTACTGGCC	297		
RESULT 5	AC005630	137699 bp	DNA	PRI 15-JAN-1999
LOCUS	AC005630			
DEFINITION	Homo sapiens clone DJ1129D05, complete sequence.			
ACCESSION	AC005630			
NID	94159882			
VERSION	AC005630.1	GI:4159882		
KEYWORDS	HTG.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 137699)			
JOURNAL	Waterston, R.H.			
REFERENCE	The sequence of Homo sapiens clone			
AUTHORS	Unpublished			
TITLE	2 (bases 1 to 137699)			
JOURNAL	Waterston, R.H.			
REFERENCE	Submitted (08-SEP-1998) Genome Sequencing Center, Washington			
AUTHORS	University School of Medicine, 4444 Forest Park Parkway, St. Louis,			
TITLE	MO 63108, USA			
JOURNAL	3 (bases 1 to 137699)			
REFERENCE	Waterston, R.H.			
AUTHORS	Direct Submission			
TITLE	Submitted (15-JAN-1999) Genome Sequencing Center, Washington			
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis,			
COMMENT	MO 63108, USA			
FEATURES	On Jan 15, 1999 this sequence version replaced gi:3550965.			
SOURCE	Location/Qualifiers			
	1. 137699			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="DJ1129D05"			
BASE COUNT	36908 a 30670 c 32412 g 37709 t			
ORIGIN				
Query Match	58.8%	Score 181.2	DB 11	Length 137699
Best Local Similarity	80.3%	Pred. No. 3e-29		
Matches 249	Conservative 0	Mismatches 58	Indels 3	Gaps 3
QY 1	CGCCGCGCACATAGCT-CCGCGCAAGTCTCGGAGCCCTTGCCATTTTCCAGCGCG-59			
DB 47458	CAGCGCGCACAGCTCTCCCGCAGAGGCTCTGGTGCCCTTGCCATTTTCCAGCGCTAC	47517		
QY 59	TCCACAGAGGTCTACGGCGCGGGGAGAGTGGAGCGCGGAGAGCTCGGCGGGGGCC	118		
DB 47518	TCCGACATAGTTCAGGAGCTACAGGAGAGCGGAGCTGGGAGAGCGCGCGAGAGGTTC	47577		
QY 119	GCCTGTGGCGCGGCATGACAGCGGCTCGGAGACTGCTCTTTTCGGCGGCCCTTCCG	178		

Db	30345	TGTGACTAGAGTTAGAGCAGCAGGGGAGAGGGCGGAGCTGGGAGACACC	CCCGGAGAGTCC	30404
QY	119	GCCGTGTGGCCCGCCCATGACACAGCGGCTGGGAGCTGCTCTTTTCGCGGCCCTCCG	178	
Db	30405	CGCGGGGTGTGGCCGCGGTACACAGCGGCTCCCGAGCGGCTCACTT	-CGGCGCCACTGTG	30463
QY	179	CCGAGGTGAGGGGAGATGTTCATGTGCAGGGTTCAAGGCCCAACGAACTTACTGGCC	238	
Db	30464	CCAGAGGTGAGATTAATATTCGTGTGAGGGTTCAAGGCCCAAGCTGAAAGTTGTTG	-AC	30522
QY	239	TCATCTTCCAGGAGAACACGAGACACAGCCGCGGGCTCAAGGCCACCGCAACATTAAG	298	
Db	30523	TCATCTTCCACAAAGAACGAGAGCCGCTCCCGAGCTCATGTCTCAGCTCAACATCAGC	30582	
QY	299	GTCAGTCGCC	308	
Db	30583	TTGAGCGCC	30592	
RESULT	7			
AC006328				
LOCUS	AC006328	178307 bp	DNA	HTG
DEFINITION	Homo sapiens clone NH0102005, WORKING DRAFT	SEQUENCE, 3 unordered		
pieces.				
AC006328				
NID	94415969			
VERSION	AC006328.2	GI:4415969		
KEYWORDS	HTG; HTGS_PHASE1.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
AUTHORS	Eutheria; Primates; Catarrhini; Homidae; Homo.			
TITLE	1 (bases 1 to 178307)			
JOURNAL	The sequence of Homo sapiens clone			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 178307)			
TITLE	Waterston, R.H.			
JOURNAL	Submitted (09-JAN-1999) Genome Sequencing Center, Washington			
REFERENCE	University School of Medicine, 444 Forest Park Parkway, St. Louis,			
AUTHORS	MO 63108, USA			
JOURNAL	On Mar 13, 1999 this sequence version replaced gi:4138780.			
COMMENT	* NOTE: This is a 'working draft' sequence. It currently			
	* consists of 3 contigs. The true order of the pieces			
	* is not known and their order in this sequence record is			
	* arbitrary. Gaps between the contigs are represented as			
	* runs of N, but the exact sizes of the gaps are unknown.			
	* This record will be updated with the finished sequence			
	* as soon as it is available and the accession number will			
	* be preserved.			
FEATURES				
source	1. 1163: contig of 1163 bp in length			
	* 1164 1181: gap of unknown length			
	* 1182 57065: contig of 45885 bp in length			
	* 57067 57084: gap of unknown length			
	* 57085 178307: contig of 12123 bp in length.			
	Location/Qualifiers			
	1. 178307			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="NH0102005"			
BASE COUNT	49986 a 38822 c 39355 g 50108 t 36 others			
ORIGIN				
Query Match	47.5%; Score 146.4; DB 35; Length 178307;			
Best local Similarity	75.5%; Ptd. No. 4, 5e-42;			
Matches 234; Conservative	0; Mismatches 71; Indels 5; Gaps 4;			
QY	1	CAGGCGCCACATAGCT-CCCGCAAGTCTCGTGGCCCTTGGCATTTCCAGCCCGC	-59	
Db	10285	CAGGCGCCACACAGAGCTCCCGCAGAGGCTTGTGTGCTTGTGCATTTTCCAGCCCTAC	10344	

```

OY 59 TCCACGAGGAGTACGCGGCGGAGAGGTGGAGCGCGGAGAGCTCGCGGCGGCGGCGG 118
Db 10345 TCGACTAGAGTGTAGGACGACGAGGAGAGCGGAGCTGGAGAGACCGCGGAGAGGTCC 10404
OY 119 GCGTGTGCGCGCGCGCATGACAGCGGCTCGGACTGCTCTTTTCGCGCGGCGGCTCCG 178
Db 10405 CCGGGGTTGTTGGCGCGCTGACAGCGGCTCCG--ACGGGCTACACTTCGCGCGGCGGCTG 10462
OY 179 CCGAGAGTGAAGGAGAGATGTCATGTCAGAGGTTCAAGGCCAAACCGAAGTACTGCGCC 238
Db 10463 CCAAGAGTGAAGTAAAGTCCGTTGAGAGGTTCAAGGCCAAAGCTAAGTGTGTGG-AC 10521
OY 239 TCAATCTCCAGAGAACGAGAGGACGACCGCGGCTCAGCGCGCGGCGGCGGCGGCGG 298
Db 10522 TCAATCTCCAGAGAACGAGAGGACGACCGCGGCTCAGCTCCTCAGCTCAGACATCAG 10581
OY 299 GTGAGTCGCC 308
Db 10582 TTGAGGCGCC 10591

JUL 8
HSU08191/c 5281 bp mRNA PRI 02-MAY-1994
LOCUS Human R kappa B mRNA, complete cds.
DEFINITION U08191.579520
ACCESSION NID 9476273
VERSION 008191.1 GI:476273
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 5281)
AUTHORS Adams,B.S., Leung,K.Y., Hanley,E.W. and Nabel,G.J.
TITLE Cloning of R kappa B, a novel DNA-binding protein that recognizes
the interleukin-2 receptor alpha chain kappa B site
New Biol. 5, 1063-1073 (1991)
JOURNAL 9215142
MEDLINE 2 (bases 1 to 5281)
REFERENCE K. Cheek.
AUTHORS Direct Submission
TITLE Submitted (25-JAN-1994) Kevin Cheek, University of Michigan Medical
JOURNAL Center, 1150 W. Medical Center Drive, Ann Arbor, MI 48109-0652, USA
COMMENT On May 2, 1994 this sequence version replaced g1:244406.
FEATURES
source
CDS

```

```

TAPITKGNLGNLSGLGNLITLTPACTKLIAGNKPVSFLTAQQLQQLQCGATQV
RIQYVPSISREEDLAPKOSPLIL"
BASE COUNT 1272 a 1508 c 1336 g 1165 t
ORIGIN
Query Match 28.4%; Score 87.4; DB 9; Length 5281;
Best Local Similarity 66.8%; Pred. No. 1.3e-09;
Matches 157; Conservative 0; Mismatches 71; Indels 7; Gaps 2;
OY 31 CGGTGCGCGCTTGCAATTTTCAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 90
Db 1098 CCGCGCGCGCTTGCAATTTTCAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1044
OY 91 GAGCCGCGAGAGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 150
Db 1044 GCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 985
OY 151 GACTGCTCTCTTTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 210
Db 984 GCCCGGCTCTCTTTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 925
OY 211 TTCAGGCGCAACGAGTACTGAGGCGGCTCTATCTTCAGAGGAGACGAGAGCGCA 265
Db 924 CTCAGGCGCGAGCTGAAGTCTCTG-CGTCACTCTTCGACAGAGACGAGAGCGCA 871

RESULT 9
OSN1A112/c 1700 bp DNA PLN 07-APR-1993
LOCUS Rice nial gene for nitrate reductase apoenzyme (exon 1-2).
DEFINITION X15819
ACCESSION NID 920270
VERSION X15819.1 GI:20270
KEYWORDS Nial gene; nitrate reductase.
SOURCE Oryza sativa.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
REFERENCE 1 (bases 1 to 1700)
AUTHORS An,G.
TITLE Direct Submission
JOURNAL Submitted (13-JUL-1989) An G., Washington State University,
PULLMAN, Washington 99164-6340, U S A
REFERENCE 2 (bases 1 to 1700)
AUTHORS Choi,H.K., Kleinhofs,A. and An,G.H.
TITLE Nucleotide sequence of rice nitrate reductase genes
JOURNAL Plant Mol. Biol. 13 (6), 731-733 (1989)
MEDLINE 91370859
COMMENT For nial
FEATURES
source
location/Qualifiers
1. 1700
/organism="Oryza sativa"
/strain="Japonica M201"
/db_xref="taxon:4530"
77. 84
/CAAT_signal
TATA_signal
CDS

```

exon 288..1326  
/number=1  
Intron 1327..1434  
/number=1  
exon 1435..1575  
/number=2  
Intron 1576..>1700  
/number=2  
BASE COUNT 320 a 578 c 494 g 308 t  
ORIGIN

Query Match 16.6%; Score 51; DB 7; Length 1700;  
Best Local Similarity 52.6%; Pred. No. 0.052;  
Matches 11; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

7 GCCACATAGCTCCGCCAAGTCTCGTGCCTTCCTTCATTTCCAGCCGCTCCACGA 66  
|||||  
533 GCCGTACAGCTCCGCCAGTCTCTGCTCCTCCTCTGCTGTCGACCGCTCGTGG 474  
|||||  
QY 67 GGGTACGCGCGGGGGAAGTGGAGCCGCGAGAGCTGCCGGGGGCCCGCTGGTG 126  
|||||  
Db 473 CGGCTTCCGCGGGGGTGGGAGAGCGAGAGGAGCCGCGACCGCGCGCGGAGTC 414  
|||||  
QY 127 GCCGCGCATGACAGCGGCTCGGAGCTGCTCTTTCGCGGGCCCTCCCGGAGGT 186  
|||||  
Db 413 GCGCGCGCTGGGATCCCTTCGCGGAGGATACGCTTGGCGCTTGAGAGAGCGGC 354  
|||||  
QY 187 GAGGGAGAGATGTCATGTCAGAGGTTCAAGG 217  
|||||  
Db 353 GCGCGCACCGCGCGGAGCCGCGCTCGAGG 323  
|||||

RESULT 10  
HS165H1F 171 bp DNA PRI 18-OCT-1995  
LOCUS H.sapiens Cpg Island DNA genomic MseI fragment, clone 165h1.  
DEFINITION forward read cpj165h1.ftla.  
ACCESSION 257156  
NID 91028387  
VERSION 257156.1 GI:1028387  
KEYWORDS Cpg Island; genomic MseI fragment.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
1 (bases 1 to 171)  
AUTHORS Primates; Catarrhini; Homnidae; Homo.  
TITLE Dodsworth,S.J., Huckle,E., Wilkinson,P. and Micklem,G.  
JOURNAL Direct Submission  
Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1RQ, England. E-mail contact: humquerry@sanger.ac.uk  
2 (bases 1 to 171)  
REFERENCE Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.  
AUTHORS Purification of Cpg Islands using a methylated DNA binding column  
JOURNAL Nat. Genet. 6 (3), 236-244 (1994)  
MEDLINE 94282070  
COMMENT Vector: pGEM-52f(-)  
Clones are available from the UK MRC Human Genome Mapping Project  
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:  
http://www.hgmp.mrc.ac.uk/ for details  
or contact: biolhelphgmp.mrc.ac.uk.  
Location/Qualifiers  
1..171

FEATURES  
Source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/sex="male"  
/dev\_stage="adult"  
/tissue\_type="blood"  
/clone\_lib="CGI-1"  
/clone="165h1"

BASE COUNT 30 a 56 c 57 g 27 t 1 others  
ORIGIN

Query Match 16.4%; Score 50.4; DB 9; Length 171;  
Best Local Similarity 85.9%; Pred. No. 0.1;  
Matches 79; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

QY 1 CAGGCCGCCATAGCTCCGCCAAGTCTCGTGCCTTCCTTCATTTCCAGCCGCG-T 59  
|||||  
Db 67 CAGGCCGCCACAGCTCCGCCAAGACTT-GGTGCCCTTGCATTTCCAGCCGCGCT 125  
|||||  
QY 60 CCCACGAGGTACAGCGCGCGGAGAGGTG 91  
|||||  
Db 126 CCCACGAGGTGAGAGCTTCAGAGAGAGCGG 157  
|||||

RESULT 11  
HS115G2F 180 bp DNA PRI 17-OCT-1995  
LOCUS H.sapiens Cpg Island DNA genomic MseI fragment, clone 115g2.  
DEFINITION forward read cpj115g2.ftla.  
ACCESSION 254415  
NID 91020456  
VERSION 254415.1 GI:1020456  
KEYWORDS Cpg Island; genomic MseI fragment.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
1 (bases 1 to 180)  
AUTHORS Primates; Catarrhini; Homnidae; Homo.  
TITLE Macdonald,M., Huckle,E., Wilkinson,P. and Micklem,G.  
JOURNAL Direct Submission  
Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1RQ, England. E-mail contact: humquerry@sanger.ac.uk  
2 (bases 1 to 180)  
REFERENCE Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.  
AUTHORS Purification of Cpg Islands using a methylated DNA binding column  
JOURNAL Nat. Genet. 6 (3), 236-244 (1994)  
MEDLINE 94282070  
COMMENT Vector: pGEM-52f(-)  
Clones are available from the UK MRC Human Genome Mapping Project  
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:  
http://www.hgmp.mrc.ac.uk/ for details  
or contact: biolhelphgmp.mrc.ac.uk.  
Location/Qualifiers  
1..180

FEATURES  
Source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/sex="male"  
/dev\_stage="adult"  
/tissue\_type="blood"  
/clone\_lib="CGI-1"  
/clone="115g2"

BASE COUNT 33 a 57 c 58 g 29 t 3 others  
ORIGIN

Query Match 16.2%; Score 49.8; DB 9; Length 180;  
Best Local Similarity 83.2%; Pred. No. 0.14;  
Matches 79; Conservative 0; Mismatches 14; Indels 2; Gaps 2;

QY 1 CAGGCCGCCATAGCTCCGCCAAGTCTCGTGCCTTCCTTCATTTCCAGCCGCG-T 59  
|||||  
Db 67 CAGGCCGCCACAGCTCCGCCAAGACTT-GGTGCCCTTGCATTTCCAGCCGCGCT 125  
|||||  
QY 60 CCCACGAGGTACAGCGCGCGGAGAGGTGAGC 94  
|||||  
Db 126 CCCACGAGGTGAGAGCTTCAGAGAGAGCGGTC 160  
|||||

RESULT 12  
HS1D8R 189 bp DNA PRI 22-OCT-1995  
LOCUS H.sapiens Cpg Island DNA genomic MseI fragment, clone 51d8, reverse  
DEFINITION

read cp951d8.rtl.a.  
 ACCESSION 261392  
 MID 91033770  
 VERSION 261392.1 GI:1033770  
 KEYWORDS Cpg island; genomic MseI fragment.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 189)  
 Macdonald, M., Huckle, E., Wilkinson, P. and Micklem, G.  
 REFERENCE Direct Submission  
 TITLE Submitted (16-Oct-1995) The Sanger Centre, Hinxton, Cambridgeshire,  
 JOURNAL CB10 1RQ, England. E-mail contact: humquerry@sanger.ac.uk  
 2 (bases 1 to 189)  
 Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.  
 REFERENCE Purification of Cpg islands using a methylated DNA binding column  
 AUTHORS TITLE Nat. Genet. 6 (3), 236-244 (1994)  
 JOURNAL MEDLINE 94282070  
 VECTOR: pGEM-5Zf(-)  
 Clones are available from the UK MRC Human Genome Mapping Project  
 Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:  
 http://www.hmp.mrc.ac.uk/ for details  
 or contact: bihelp@hmp.mrc.ac.uk.  
 Location/Qualifiers  
 1. 189  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /sex="male"  
 /dev\_stage="adult"  
 /tissue\_type="blood"  
 /clone\_lib="CGI-1"  
 /clone="51d8"  
 /clone="51d8" 61 c 71 g 22 t 3 others  
 BASE COUNT 32 a 61 c 71 g 22 t 3 others  
 ORIGIN  
 Query Match 16.0%; Score 49.4; DB 9; Length 189;  
 Best Local Similarity 84.8%; Pred. No. 0.16;  
 Matches 67; Conservative 0; Mismatches 11; Indels 1; Gaps 1;  
 QY 215 AGCCCAACGAGTACTGCGCCCTATCTTCACAGAGACGACGACCGCGG 274  
 DB 1 AGCCCAACGAGTACTGCGCCCTATCTTCACAGAGACGACGACCGCGG 59  
 QY 275 CTCACGCCCGACGCGCAACA 293  
 60 CTCACGCCCGACGCGCAACA 78  
 RESULT 13  
 AC004690/c 142123 bp DNA PRI 16-MAY-1998  
 LOCUS Homo sapiens PAC clone DJ0630C24 from 7q31-q32, complete sequence.  
 DEFINITION AC004690  
 ACCESSION 93135287  
 MID 93135287  
 VERSION AC004690.1 GI:3135287  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 142123)  
 Kallio, J. and Burkhardt, J.  
 REFERENCE Unpublished (1998)  
 TITLE The sequence of Homo sapiens PAC clone DJ0630C24  
 JOURNAL 2 (bases 1 to 142123)  
 REFERENCE Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-MAY-1998) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT SUBMITTED BY: W05SC  
 Genome Sequencing Center

Department of Genetics  
 Washington University  
 St. Louis MO 63108, USA  
 http://genome.wustl.edu/gsc  
 mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this  
 clone. It may be shorter because we only sequence overlapping  
 clone sections once, or longer because we provide a small overlap  
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded or sequenced with an alternate  
 chemistry; an attempt was made to resolve all sequencing problems,  
 such as compressions and repeats; all regions were covered by  
 sequence from more than one subclone; and the assembly was  
 confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and  
 sequencing collaboration between the NHGRI Chromosome 7 Mapping  
 Project (Eric D. Green, Director), John D. McPherson in the  
 Department of Genetics (Washington University), and the Washington  
 University Genome Sequencing Center. For additional information  
 about the map position of this sequence, see  
 http://www.nhgri.nih.gov/DIR/GRB/CHR/, send  
 mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

#### SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-4, prepared by  
 Pieter de Jong and coworkers at Roswell Park Cancer Institute,  
 using the method described by Roanann et al., Nature Genetics,  
 6:84-9 (1994). The library is from one male donor. For further  
 details, see http://bacpac.med.buffalo.edu/  
 The clone is available from Genome Systems, Inc.  
 (http://www.genomesystems.com).

#### VECTOR: RPCI-4

NEIGHBORING SEQUENCE INFORMATION:  
 The actual start of this clone is at base position 1 of DJ0630C24;  
 actual end is at 142123 of DJ0630C24.

This clone contains STS SWS3567 (NID:q1916499).

This clone contains a transposon, from base 123435-124772, which  
 has been deleted from the submitted sequence.

#### FEATURES

##### source

1. 142123

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="7"

/map="7q31-q32"

/clone\_lib="RPCI-4"

/clone="DJ0630C24"

1. 95

/rpt\_family="Alu"

/rpt\_family="MIR"

2378..2865

/rpt\_family="MER50"

2964..3027

/rpt\_family="MER50"

3463..3660

/rpt\_family="Alu"

3735..3765

/rpt\_family="L1"

3766..3858

/rpt\_family="L1"

4018..4551

/rpt\_family="L1"

7175..7409

/rpt\_family="MER2\_type"

8426..8775

/rpt\_family="MIR"

```
repeat_region /rpt_family="Alu" 45759..45881
repeat_region /rpt_family="L1" 45902..46280
repeat_region /rpt_family="L1" 48272..48501
repeat_region /rpt_family="L1" 48505..49036
repeat_region /rpt_family="L1" 49052..49652
repeat_region /rpt_family="L1" 49647..50340
repeat_region /rpt_family="MaltR" 50341..50720
repeat_region /rpt_family="MaltR" 50723..51658
repeat_region /rpt_family="L1" 52203..52314
repeat_region /rpt_family="L2" 52505..52616
repeat_region /rpt_family="MER1_type" 52617..52789
repeat_region /rpt_family="Alu" 52790..52873
repeat_region /rpt_family="MER1_type" 52937..53912
repeat_region /rpt_family="L2" 54049..54356
repeat_region /rpt_family="Alu" 54511..55100
repeat_region /rpt_family="Retroviral" 56013..56227
repeat_region /rpt_family="MIR"
```

Query Match	14.6%	Score 45;	DB 11;	Length 142123;
Best Local Similarity	53.8%	Fred. No. 0.42;		
Matches 93;	Conservative	0;	Mismatches 80;	Indels 0;
				Gaps 0;
Qy 18	CCCGCCAGATCTCTCGATCGATCCCTTCCATTTTCCAGCCGCGCTCCACAGAGGGGTCAACGGCG 77			
Db 18251	CCCGCCATGGGGCTCGGAGAGCTGTCCGAACGCCCGCAGCTGCACACCCCGCGCTGCGGGGT 18192			
Qy 78	GCGGCGAGAGGTGTGACGCCGCGAGAGCTCGGCCCGGGGGCCCGCTGTGTGGCCGCGCCCAT 137			
Db 18191	GCGGCGAGACTCAGAGGGCCGAGGGGCCGAGACCCCGGCGCGCTCTGTGGCCGCTGCGGGCCC 18132			
Qy 138	GACAGCGGCTGTGGGACTGGCTTCCTTTTCCGGCCCCCTCCCGCGGAGGTGAG 190			
Db 18131	CCGGGGGCGCGCGACAGAGGAGCTCCGTCGCCGCCGCGCTCCGGCGGAGCGGGGG 18079			
RESULT 14				
HSCG16P5/c				
LOCUS	HSCG16P5. 2837 bp DNA			
DEFINITION	Homo sapiens DNA for (CGG)n trinucleotide repeat region, isolate			
ACCESSION	AF001218.1 GI:3790170			
NID	93790170			
VERSION	trinnucleotide repeat.			
KEYWORDS	human.			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;			
REFERENCE	Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 2837)			
AUTHORS	Mangel, L.			
JOURNAL	Direct Submission			
	Submitted (12-AUG-1997) Mangel L., Medical Genetics and Virology,			
	University of Cologne, Institute of Genetics, Weyerlall 121,			
	Cologne, 50931, GERMANY			
REMARK	Revised by [3]			
REFERENCE	2 (bases 1 to 2837)			
AUTHORS	Mangel, L.			

[illegible]

Sun Sep 19 10:27:09 1999

us-08-956-518a-95.rge

Page 10

Db 3633 GCACCGCTCGAGCGCTCGACCGGTCTCGCCGGGGGGGTGGGGCCGAGCCGCGCTAC 3692  
OY 233 TGGCCCTTATCTTCAGAGAGACCAGGAGCCACAGCCGCGGCTCAGCCGCCACCGCA 290  
Db 3693 CGCGACCCGTCGCGGGGGCGAGCTGTGGGCGTGCAGCTGGGCCAGGCGCAGCAGCA 3750

Search completed: September 17, 1999, 22:03:03  
Job time: 16366 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:10 ; Search time 2825.05 seconds  
(without alignments)  
215.055 Million cell updates/sec

Title: US-08-956-518a-95

Perfect score: 308  
Sequence: 1 CAGGCCGCCACATAGCTCC.....CAACATTAGGTAGTCGCC 308

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 98626752 residues

base :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: em\_est20: \*  
21: em\_est21: \*  
22: em\_est22: \*  
23: em\_est23: \*  
24: em\_est24: \*  
25: em\_est25: \*  
26: em\_est26: \*  
27: em\_est27: \*  
28: em\_est28: \*  
29: em\_est29: \*  
30: em\_est30: \*  
31: em\_est31: \*  
32: em\_est32: \*  
33: em\_est33: \*  
34: em\_est34: \*  
35: em\_est35: \*  
36: em\_est36: \*  
37: em\_est37: \*  
38: em\_est38: \*  
39: em\_est39: \*  
40: em\_est40: \*  
41: em\_est41: \*  
42: em\_est42: \*  
43: em\_est43: \*  
44: em\_est44: \*  
45: em\_est45: \*  
46: em\_est46: \*  
47: em\_est47: \*  
48: em\_est48: \*  
49: em\_est49: \*  
50: em\_est50: \*  
51: em\_est51: \*  
52: em\_est52: \*  
53: em\_est53: \*

54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	93.2	30.3	170	25	N83152	N83152 K4329F Huma
2	81.2	26.4	510	54	HSN011948	A1047098 Homo Sap
3	57.6	18.7	335	32	AA333244	AA333244 EST0317
4	43.8	14.2	416	33	AA407437	AA407437 EST00734
5	42	13.6	328	33	AA407004	AA407004 EST02243
6	41.6	13.5	213	33	AA410084	AA410084 EST01971
7	40.4	13.1	268	33	AA407006	AA407006 EST02245
8	40.4	13.1	229	33	AA410099	AA410099 EST01951
9	40.4	13.1	325	44	A1317247	A1317247 u136c11.Y
10	40	13.0	371	47	A1508264	A1508264 mb6603.Y
11	39.8	12.9	374	33	AA407075	AA407075 EST02080
12	39.8	12.9	376	33	AA407297	AA407297 EST02294
13	39.8	12.9	271	33	AA407662	AA407662 EST01743
14	39.8	12.9	316	33	AA407690	AA407690 EST01973
15	39.8	12.9	373	33	AA407899	AA407899 EST01001
16	39.8	12.9	322	33	AA407912	AA407912 EST01757
17	39.8	12.9	224	33	AA407943	AA407943 EST01023
18	39.8	12.9	335	33	AA408014	AA408014 EST02534
19	39.8	12.9	388	33	AA408056	AA408056 EST01037
20	39.8	12.9	354	33	AA408063	AA408063 EST01129
21	39.8	12.9	376	33	AA408103	AA408103 EST02583
22	39.8	12.9	321	33	AA408119	AA408119 EST02666
23	39.8	12.9	261	33	AA408198	AA408198 EST02863
24	39.8	12.9	321	33	AA408495	AA408495 EST02965
25	39.8	12.9	437	33	AA408575	AA408575 EST03109
26	39.8	12.9	485	33	AA408661	AA408661 EST03176
27	39.8	12.9	414	33	AA408707	AA408707 EST03168
28	39.8	12.9	525	33	AA408911	AA408911 EST03425
29	39.8	12.9	308	33	AA409494	AA409494 EST01256
30	39.8	12.9	315	33	AA410085	AA410085 EST01972
31	39.8	12.9	372	45	A1324211	A1324211 m122c07.X
32	39.8	12.9	334	45	A1324741	A1324741 mb78c09.X
33	39.8	12.9	373	45	A1324799	A1324799 mc31n01.X
34	39.8	12.9	356	46	A1414218	A1414218 mb61e11.X
35	39.8	12.9	365	46	A1414287	A1414287 mb81f02.X
36	39.8	12.9	372	46	A1426055	A1426055 mg53a11.X
37	39.8	12.9	335	46	A1430785	A1430785 mb66d03.X
38	39.8	12.9	372	46	A1449591	A1449591 mq06e08.X
39	39.8	12.9	354	46	A1452165	A1452165 mp81e11.X
40	39.8	12.9	372	47	A1463228	A1463228 uc39b02.X
41	39.8	12.9	372	50	A1666386	A1666386 mu14a08.X
42	38.6	12.5	369	29	AA156428	AA156428 z151c02.S
43	38.4	12.5	312	33	AA410147	AA410147 EST01892
44	38.4	12.5	372	49	A1645460	A1645460 v188b01.X
45	38.2	12.4	354	46	A1426052	A1426052 mg52e10.X

#### ALIGNMENTS

RESULT 1  
N83152  
LOCUS N83152 170 bp mRNA  
DEFINITION K4329F Human fetal heart, Lambda ZAP Express Homo sapiens CDNA  
ACCESSION N83152  
NID g1441866  
VERSION N83152.1 GI:1441866



Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Palligano, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Danke, D., Feng, D.-F., Ferris, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meisner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wang, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M., and Venter, J.C.

Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

On May 18, 1995 this sequence version replaced g1:811121.

Other ESTs: THC185696

Contact: Kerlavage, AR

Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org

For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi.html>)

Seq primer: M13 Reverse.

Location/Qualifiers

1..335

/organism="Homo sapiens"

/db\_xref="ATCC (inhost):134920"

/db\_xref="taxon:9606"

/clone\_lib="Embryo, 8 week 1"

/dev\_stage="Embryo, 8 wks"

/note="Organ: Embryo, 8 weeks; Vector: pBluescript SK-; Site\_1: EORI; Site\_2: XhoI"

Site\_1: EORI; Site\_2: XhoI

84 a 94 c 88 g 69 t

BASE COUNT

ORIGIN

Query Match 18.7%; Score 57.6; DB 32; Length 335;  
Best Local Similarity 83.7%; Pred. No. 0.00069;  
Matches 77; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 174 TCCGCGCGGAGTGGAGGAGATGTCATGTCAGGCTCAAGCCAAACGAGTACT 233  
1 TCCGCGCGGAGTGGAGGAGATGTCATGTCAGGCTCAAGCCAAACGAGTACT 60

QY 234 GGCCCTCTATCTTCAGAGACCAAGCAGCA 265  
1 GGCCCTCTATCTTCAGAGACCAAGCAGCA 91

Db 61 GG-CGTCCATCTTCAGACAGACCAAGCAGCA 91

RESULT 4  
AA407437 416 bp mRNA EST 26-AUG-1998  
LOCUS AA407437  
DEFINITION EST00734 Mouse 7.5 dpc embryo ectoplacental cone cDNA library Mus musculus cDNA clone C0002B08 3', mRNA sequence.  
ACCESSION AA407437  
NID 92065649  
VERSION AA407437.1 GI:2065649  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 416)  
Ko, M.S.H., Threat, T.A., Horton, J.H., Wang, X., Cui, Y., Wang, X., Pryor, E., Paris, J., Wells-Smith, J., Fujiwara, H., Yotsumoto, S. and Nakashima, H.  
Systematic analyses of mouse genes expressed in embryo implantation site  
Unpublished (1997)

On Sep 12, 1996 this sequence version replaced g1:1292065.

Contact: KO MSH  
Center for Molecular Medicine and Genetics  
Wayne State University  
5047 Guillen Mall, Detroit, MI 48202  
Tel: 3135776708  
Fax: 3135776200  
Email: mskoe@emb.biosci.wayne.edu  
Human 28S ribosomal RNA gene.

Seq primer: M13 Forward.

Location/Qualifiers

1..416

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="ATCC (inhost):1363489"

/db\_xref="taxon:10090"

/map="20"

/clone="C0002B08"

/clone\_lib="Mouse 7.5 dpc embryo ectoplacental cone cDNA library"

/sex="unknown"

/tissue\_type="ectoplacental cone"

/dev\_stage="embryonic day 7.5 postconception"

/lab\_host="DH10B"

/note="Organ: embryo; Vector: pSPORT1 (Life Technologies); Site\_1: SalI; Site\_2: NotI; Total RNAs were extracted from ectoplacental cone of 7.5-dpc embryos. The double-stranded cDNA was synthesized from total RNAs with an Oligo(dT) primer. The library was constructed by Minoru S. H. Ko."

S. H. Ko."

67 a 130 c 133 g 69 t 17 others

BASE COUNT

ORIGIN

Query Match 14.2%; Score 43.8; DB 33; Length 416;  
Best Local Similarity 54.2%; Pred. No. 0.73;  
Matches 84; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 28 CCTCGTGCCCTTTCAGATTTCCAGCCGCGTCCAGAGAGGTACAGCGCGGAGAG 87  
1 CCTCGTGCCCTTTCAGATTTCCAGCCGCGTCCAGAGAGGTACAGCGCGGAGAG 100

Db 41 CCGCGGGCGCNCNGCNCACCCCGCCGCTCTCCGAGGGGCGGCGGCGCGC 100

QY 88 GTGAGCCCGGAGAGCTCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 147  
1 GTGAGCCCGGAGAGCTCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 160

Db 101 GGCGTCCCGCGGTGTGGGAGACCTCCGCTGCTTCCCGCGCGCGCGCGCGCGCC 160

QY 148 CGGGAGTGGCTCTTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 182  
1 CGGGAGTGGCTCTTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 195

Db 161 CGGGCGCGGTTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 195

RESULT 5  
AA407004 328 bp mRNA EST 26-AUG-1998  
LOCUS AA407004  
DEFINITION EST02243 Mouse 7.5 dpc embryo ectoplacental cone cDNA library Mus musculus cDNA clone C0019B05 3', mRNA sequence.  
ACCESSION AA407004  
NID 92067158  
VERSION AA407004.1 GI:2067158  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 328)  
Ko, M.S.H., Threat, T.A., Horton, J.H., Wang, X., Cui, Y., Wang, X., Pryor, E., Paris, J., Wells-Smith, J., Fujiwara, H., Yotsumoto, S. and Nakashima, H.  
Systematic analyses of mouse genes expressed in embryo implantation site  
Unpublished (1997)

JOURNAL

COMMENT On Sep 12, 1996 this sequence version replaced g1:1393284.

—



Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:968824  
 Seq primer: -40RP from Glibco.  
 Location/Qualifiers

# FEATURES

## SOURCE

1.325  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /map="12, 12, 40, 8CR from WI-4263; 21q"  
 /clone="IMAGE:1853396"  
 /sex="equal ratio of male:female"  
 /tissue="urogenital ridge (embryonic)"  
 /dev\_stage="fetal, mixture of 11.5 and 12.5 dpc"  
 /lab\_host="DH10B"  
 /note="Organ: gonad; Vector: pT73D-Pac (Pharmacia) with a  
 modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5',  
 TGTACCAATCTGAAGTGGAGCGCGCGGCAATTTTCTTTTCTTTTCTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT73 vector.  
 Library went through two rounds of normalization, and was  
 constructed by Bento Soares and M.Fatima Bonaldo."

## BASE COUNT

54 a 114 c 107 g 49 t 1 others

## ORIGIN

### Query Match

Best Local Similarity 13.1%; Score 40.4; DB 44; Length 325;  
 Matches 83; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 28 CCGCGTCCCTTCCATTTCCAGCCGCGTCCACGAGGTCAGCGCGGAGAG 87  
 DB 324 CCGCGGCGCTCTCCACCCCGGCTCTCCGAGGTCAGCGCGGCGGCG 265

QY 88 GTGAGCGCGGAGAGTCCGCGCGGCGCGCTGTCGCGCGGCGGAGAG 147  
 DB 264 GGGGTCCCGCGGTGTGGGGGGAACCTCCGCGGTTCGCCGCCGCGGCG 205

QY 148 CGGAGTGGCTCTTTCGCGCGCGGCGCGGCGGCGGCGGCGGCG 182  
 DB 204 CGGCGCGGCTTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCG 170

## RESULT 10

### DEFINITION

A1508264 371 bp mRNA EST 12-MAR-1999  
 mh6d03.y1 Soares mouse placenta 4NBP13.5 14.5 Mus musculus cDNA  
 clone IMAGE:455909 5' similar to TR:004892 004892 CYTOCHROME P450  
 Like TSP, mRNA sequence.

## ACCESSION

A1508264

## NID

94407169

## VERSION

A1508264.1 GI:4407169

## KEYWORDS

EST.

## SOURCE

house mouse.

## ORGANISM

Mus musculus

## REFERENCE

Authors

## TITLE

Journal

## COMMENT

On May 18, 1998 this sequence version replaced gi:3138778.

Contact: Marra M/Mashu-NCI Mouse EST Project 1999  
 Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 This read is a RESEQUENCE of a previously sequenced mouse clone  
 correct orientation)  
 Seq primer: -40RP from Glibco  
 High quality sequence stop: 337.  
 Location/Qualifiers

# FEATURES

## SOURCE

1.371  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /map="15 q22-qter; 21q"  
 /clone="IMAGE:455909"  
 /clone\_lib="Soares mouse placenta 4NBP13.5 14.5"  
 /sex="unknown"  
 /tissue="placenta"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Organ: placenta; Vector: pT73D-Pac (Pharmacia)  
 with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
 1st strand cDNA was primed with a Not I - oligo(dT) primer  
 [5',  
 TGTACCAATCTGAAGTGGAGCGCGCGGCAATTTTCTTTTCTTTTCTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M.Fatima Bonaldo."

## BASE COUNT

58 a 135 c 127 g 50 t 1 others

### Query Match

Best Local Similarity 13.0%; Score 40; DB 47; Length 371;  
 Matches 85; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 22 CCAATCTCCGCTCCCTTCCATTTCCAGCCGCGTCCACGAGGTCAGCGGCGG 81  
 DB 330 CCTTCCGCGGCGGCGCTCTCCACCCCGGCTCTCCGAGGTCAGCGGCGG 271

QY 82 GGAGAGTGGAGCGCGGAGAGTGGCGCGGCGGCGGCGGCGGCGGCGGCG 141  
 DB 270 CGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 211

QY 142 GCGGCTCGGAGTGGCTCTTTCGCGCGCGGCGGCGGCGGCGGCGGCGG 182  
 DB 210 GCCCGCGGCGGCGGCTTTCGCGCGGCGGCGGCGGCGGCGGCGGCGG 170

## RESULT 11

### DEFINITION

AA407075 374 bp mRNA EST 26-AUG-1998  
 EST02080 Mouse 7.5 dpc embryo ectoplacental cone cDNA library Mus  
 musculus cDNA clone C0017C05 3', mRNA sequence.

## ACCESSION

AA407075

## NID

92066995

## VERSION

AA407075.1 GI:2066995

## KEYWORDS

EST.

## SOURCE

house mouse.

## ORGANISM

Mus musculus

## REFERENCE

Authors

## TITLE

Systematic analyses of mouse genes expressed in embryo implantation site

site  
 Systematic analyses of mouse genes expressed in embryo implantation



Center for Molecular Medicine and Genetics  
Wayne State University  
5047 Guillen Mall, Detroit, MI 48202  
Tel: 3135776708  
Fax: 3135776200

Email: mskocmb.biosci.wayne.edu

Mouse 28S ribosomal RNA

Seq primer: M13 Forward

## FEATURES

## source

Location/Qualifiers

1. 271

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="ATCC (inhost):1364262"

/db\_xref="taxon:10090"

/clone="C0013E08"

/clone\_lib="Mouse 7.5 dpc embryo ectoplacental cone cDNA library"

/sex="unknown"

/tissue\_type="ectoplacental cone"

/dev\_stage="embryonic day 7.5 postconception"

/lab\_host="DH10B"

/note="Organ: embryo; Vector: pSPORT1 (Life Technologies); Site\_1: SalI; Site\_2: NotI; Total RNAs were extracted from ectoplacental cone of 7.5-dpc embryos. The double-stranded cDNA was synthesized from total RNAs with an Oligo(dT) primer. The library was constructed by Minoru S. H. Ko."

BASE COUNT 29 a 105 c 96 g 39 t 2 others

ORIGIN

## Query Match

Best Local Similarity 12.9%; Score 39.8; DB 33; Length 271;

Matches 83; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 28 CCTGCGTCCCTTCCATTTTCACGCCGTCACAGAGGTGTCAGCGGGAGAG 87

DB 40 CCGCGGGGCTGCTGCACACCCCATGCTCTCCAGAGGTGCGGGGGCGG 99

QY 88 GTGAGCGCGGAGAGCTGCGGGGGCGCGCGTGTGCGCGGCATGACAGCGGCT 147

DB 100 GCGTCCCGCGGTGTGCGGGGGAACCTCCGCTGCTTCCCGCGGGTCCGCCCC 159

QY 148 CGGACTGCTCTTTCCTTCCGCGCCCTCCGCCGG 182

DB 160 CGGCGCGGCTTTCCTTCCGCGCGCTCCGCCGG 194

QY 148 CGGACTGCTCTTTCCTTCCGCGCCCTCCGCCGG 182

DB 160 CGGCGCGGCTTTCCTTCCGCGCGCTCCGCCGG 194

LOCUS

AA407690

DEFINITION

EST01973 Mouse 7.5 dpc embryo ectoplacental cone cDNA library Mus

musculus cDNA clone C0016C02 3', mRNA sequence.

ACCESSION

AA407690

NID

92066888

VERSION

AA407690.1

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 316)

Ko, M.S.H., Threat, T.A., Horton, J.H., Wang, X., Cul, Y., Wang, X.,

Prior, E., Paris, J., Wells-Smith, J., Fujiwara, H., Yotsumoto, S. and

Nakashima, H.

Systematic analyses of mouse genes expressed in embryo implantation

site

Unpublished (1997)

On Sep 12, 1996 this sequence version replaced gi:1392824.

Contact: Ko MSH  
Center for Molecular Medicine and Genetics  
Wayne State University

5047 Guillen Mall, Detroit, MI 48202

Tel: 3135776708

Fax: 3135776200

Email: mskocmb.biosci.wayne.edu

Human 28S ribosomal RNA gene.

Seq primer: M13 Forward.

## FEATURES

## source

Location/Qualifiers

1. 316

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="ATCC (inhost):1363618"

/db\_xref="taxon:10090"

/clone="C0016C02"

/clone\_lib="Mouse 7.5 dpc embryo ectoplacental cone cDNA library"

/sex="unknown"

/tissue\_type="ectoplacental cone"

/dev\_stage="embryonic day 7.5 postconception"

/lab\_host="DH10B"

/note="Organ: embryo; Vector: pSPORT1 (Life Technologies); Site\_1: SalI; Site\_2: NotI; Total RNAs were extracted from ectoplacental cone of 7.5-dpc embryos. The double-stranded cDNA was synthesized from total RNAs with an Oligo(dT) primer. The library was constructed by Minoru S. H. Ko."

BASE COUNT 33 a 118 c 115 g 50 t

ORIGIN

## Query Match

Best Local Similarity 12.9%; Score 39.8; DB 33; Length 316;

Matches 89; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 28 CCTGCGTCCCTTCCATTTTCACGCCGTCACAGAGGTGTCAGCGGGAGAG 87

DB 38 CCGCGGGGCTGCTGCACACCCCATGCTCTCCAGAGGTGCGGGGGCGG 97

QY 88 GTGAGCGCGGAGAGCTGCGGGGGCGCGCGTGTGCGCGGCATGACAGCGGCT 147

DB 98 GCGTCCCGCGGTGTGCGGGGGAACCTCCGCTGCTTCCCGCGGGTCCGCCCC 157

QY 148 CGGACTGCTCTTTCCTTCCGCGCCCTCCGCCGG 182

DB 158 CGGCGCGGCTTTCCTTCCGCGCGCTCCGCCGG 192

QY 148 CGGACTGCTCTTTCCTTCCGCGCCCTCCGCCGG 182

DB 158 CGGCGCGGCTTTCCTTCCGCGCGCTCCGCCGG 192

LOCUS

AA407899

DEFINITION

EST01001 Mouse 7.5 dpc embryo ectoplacental cone cDNA library Mus

musculus cDNA clone C0004G11 3', mRNA sequence.

ACCESSION

AA407899

NID

92065916

VERSION

AA407899.1

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 373)

Ko, M.S.H., Threat, T.A., Horton, J.H., Wang, X., Cul, Y., Wang, X.,

Prior, E., Paris, J., Wells-Smith, J., Fujiwara, H., Yotsumoto, S. and

Nakashima, H.

Systematic analyses of mouse genes expressed in embryo implantation

site

Unpublished (1997)

On Sep 12, 1996 this sequence version replaced gi:1324748.

Contact: Ko MSH  
Center for Molecular Medicine and Genetics  
Wayne State University  
5047 Guillen Mall, Detroit, MI 48202  
Tel: 3135776708



Fax: 3135776200

Email: mskoe@cmb.biosci.wayne.edu

Human ribosomal DNA complete repeating unit  
Seq primer: M13 Forward.

## FEATURES

source

Location/Qualifiers

1..373  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="ATCC (inhost):1364870"  
 /db\_xref="taxon:10090"  
 /clone="C0004G11"  
 /clone\_lib="Mouse 7.5 dpc embryo ectoplacental cone cDNA library"  
 /sex="unknown"  
 /tissue\_type="ectoplacental cone"  
 /dev\_stage="embryonic day 7.5 postconception"  
 /lab\_host="DH10B"  
 /note="Organ: embryo; Vector: pSPORT1 (Life Technologies); Site\_1: SalI; Site\_2: NotI; Total RNAs were extracted from ectoplacental cone of 7.5-dpc embryos. The double-stranded cDNA was synthesized from total RNAs with an Oligo(dt) primer. The library was constructed by Minoru S. H. Ko."

BASE COUNT 56 a 122 c 132 g 61 t 2 others  
 ORIGIN

## Query Match

Best Local Similarity 12.9%; Score 39.8; DB 33; Length 373;  
 Matches 83; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 28 CCTCGGTGCCCCCTTTCACAGCCGCTGCCACAGAGGTCACGCGCGGAGAG 87  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 40 CCGCGGGGCTCTCGCTCCACCCGCCATGCTCTCCGAGGTGGGTGGGGGGGGCG 99  
 QY 88 GTGAGACCGCAGACTGGCGCGGGGGCCCGCTGTGTGCCCGGCATGACAGCGGCT 147  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 100 GGCGTCCCGCGCGGTGTGGGGGGAACCTCGCGTGTCTCCCGCGGCGGTCCGCCCCC 159  
 QY 148 CGGGACTGGCTCTTTTCCGCGCCCTCCGCGG 182  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 160 CGGGCGCGGTTTTCGCGCGCGCTCGCTCGG 194

Search completed: September 17, 1999, 21:28:13  
 Time: 14277 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:03:03 ; Search time 1962.17 Seconds

(Without alignments)  
235.018 Million cell updates/sec

Title: US-08-956-518a-96

Perfect score: 145

Sequence: 1 CTCATTCAGATTACAAAGTG.....ACTCTGACAGCTGACTTGTGA 145

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

abase :

GenEmbl.\*  
1: gb\_da1:\*  
2: gb\_da2:\*  
3: gb\_cm:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_to:\*  
13: gb\_st:\*  
14: gb\_sy:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_vl:\*  
18: em\_fun:\*  
19: em\_htg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_cm:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_to:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_vl:\*  
34: gb\_htg1:\*  
35: gb\_htg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_da1:\*  
39: em\_da2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pr4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	129.2	89.1	641	11	AF029838	AF029838 Homo sapi

2	128.8	88.8	689	11	AF029839	AF029839 Homo sapi
3	34.4	23.7	3093	12	RND0CP	X13417 Rat ornithi
4	30.4	21.0	491	7	MINC01	V00667 Neurospora
5	30.2	20.8	143749	11	AC004962	AC004962 Homo sapi
6	30.0	20.7	146058	11	AC003668	AC003668 Homo sapi
7	29.8	20.6	196589	34	AF128834	AF128834 Homo sapi
8	29.8	20.6	132050	42	AC004909	AC004909 Homo sapi
9	29.4	20.3	2860	8	AF100163	AF100163 Arabidops
10	29.4	20.3	105654	8	ATAF001308	ATAF001308 Arabidops
11	29.4	20.3	73289	19	AC005981	AC005981 *** SEQUE
12	29.4	20.3	81146	35	AC005427	AC005427 Drosophi
13	29.4	20.3	62484	35	AC005453	AC005453 Drosophi
14	29.4	20.3	35807	36	CET13H5	Z6524 Caenorhabd
15	29.4	20.3	60019	37	AC005650	AC005650 Drosophi
16	29.2	20.1	100639	34	AC002417	AC002417 Homo sapi
17	28.6	19.7	2036	3	P1GSLTXX	L02900 Pig Na-depe
18	28.6	19.7	2542	7	ATC1CB	Z71446 A.thaliana
19	28.6	19.7	2755	8	AF133209	AF133209 Nicotiana
20	28.6	19.7	1354	10	HSN800075	AL050297 Homo sapi
21	28.4	19.6	12631	2	AE001282	AE001282 Chlamydia
22	28.4	19.6	175120	11	AC004687	AC004687 Homo sapi
23	28.4	19.6	6309	12	AF015311	AF015311 Rattus no
24	28.4	19.6	173616	35	AC006509	AC006509 Homo sapi
25	28	19.3	4169	7	ANSAMB	AJ000996 Aspergill
26	28	19.3	2040	7	OSCDPK2	X81394 O.saliva mr
27	27.8	19.3	102402	35	AC007641	AC007641 Mus muscu
28	27.8	19.2	100000	9	AP000068	AP000068 Homo sapi
29	27.8	19.2	165139	10	HS436K10	AL031407 Human DNA
30	27.8	19.2	273800	12	AF100956	AF100956 Mus muscu
31	27.8	19.2	213245	12	MMHC425018	AF110520 Mus muscu
32	27.8	19.2	11726	34	CET21E11	AL009107 Caenorhab
33	27.8	19.2	281947	35	AC007569	AC007569 Homo sapi
34	27.8	19.2	94718	36	AC004658	AC004658 Drosophi
35	27.8	19.2	26208	36	CEC40C9	Z70266 Caenorhabd
36	27.8	19.2	3722	36	CER42H11	Z83225 Caenorhabd
37	27.6	19.0	3781	8	AF030634	AF030634 Neurospor
38	27.4	18.9	1541	8	ATU80185	U80185 Arabidopsis
39	27.4	18.9	100000	9	AP000147	AP000147 Homo sapi
40	27.4	18.9	89086	9	AP000233	AP000233 Homo sapi
41	27.4	18.9	3340	12	AF136401	AF136401 Rattus no
42	27.4	18.9	171914	34	AC006169	AC006169 Drosophi
43	27.4	18.9	85096	35	AC007323	AC007323 Arabidops
44	27.2	18.8	7150	9	AB007931	AB007931 Homo sapi
45	18.8	36133	9	9	HSA007973	AJ007973 Homo sapi

## ALIGNMENTS

RESULT 1	AF029838	641 bp	MRNA	PRI	16-DEC-1998
LOCUS	Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.				
DEFINITION	AF029838				
ACCESSION	AF029838				
NID	93757793				
VERSION	AF029838.1	GI:3757793			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	Gault,J., Robinson,M., Berger,R., Drebing,C., Logel,J., Hopkins,J., Moore,T., Jacobs,S., Merivether,J., Choi,M.J., Kim,E.J., Walton,K., Bulding,K., Davis,A., Breese,C., Freedman,R. and Leonard,S.				
TITLE	Genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene				
JOURNAL	Genomics 52 (2), 173-185 (1998)				
MEDLINE	99000837				
REFERENCE	2 (bases 1 to 641)				
AUTHORS	Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R., Breese,C., Davis,A., Hopkins,J. and Freedman,R.				
TITLE	Direct Submission				

Submitted (15-OCT-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA

```

FEATURES
source      Location/Qualifiers
1. .641

```

```

misc.feature
<1..>641
/note="A/B/C/D"
/clone="CHRNA7-DR1"
/map="15q14"
/chromosome="15"
/db_xref="taxon:9606"
/organism="Homo sapiens"
alternative splicing"
BASE COUNT 151 a 192 c 163 g 135 t
ORIGIN

```

Query Match	89.18;	Score 129.2;	DB 11;	Length 641;
Best Local Similarity	97.88;	Pred. No. 7.1e-33;		
atches 131; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0

Oy	6	TTGAGATTACAAAGTGGGAACCTGAGATCAGAGGAACCTGGAAATCCCAAGTGTGAGAGCTTA	65
Db	294	TTAAGATTACAAAGTGGGAACCTGAGATCAGAGGAACCTGGAAATCCCAAGTGTGAGAGCTTA	35
Oy	66	TCCTACAGCACTCAGATCTTGTGTCCACCCCATTTATGACATCCAAAGTGTGAGAAAGC	12
Db	354	TCCTACAGCACTCAGATCTTGTGTCCACCCCATTTATGACATCCAAAGTGTGAGAAAGC	41
Oy	126	ACTGTGACAACTGA	139
Db	414	ACTGTGACAACTGA	427

RESULT	2
AF029839	
LOCUS	689 bp mRNA
DEFINITION	Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence
ACCESSION	AF029839
NID	G9757794
VERSION	AF029839.1 GI:3757794
KEYWORDS	
SOURCE	human.

BASE COUNT	142 a	210 c	196 g	141 t
ORIGIN				

Query Match	88.8%;	Score 128.8;	DB 11;	Length 689;
Best Local Similarity	98.5%;	Pred. No. 9.7e-33;		
Matches 130; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

QY	6	TTGAGATTACAAGTGGACACTGTAGTCAGAGAGACCTGGAAATCCGAGATCAGAGAGCTTA	65
Db	406	TTAGATTTCACAAGTGGACACCTGTAGTCAGAGAGACCTGGAAATCCGAGATCAGAGAGCTTA	465
QY	66	TCTACAGAGACTCAGATTTTGTGTGCACCCCATATTGACAAATCCAAAGTGCAGAAAGC	125
Db	466	TCTACAGAGACTCAGATTTTGTGTGCACCCCATATTGACAAATCCAAAGTGCAGAAAGC	525
QY	126	ACTCTGACAAGT	137
Db	526	ACTCTGACAAGT	537

RESULT	3
LOCUS	RNODCP/C
DEFINITION	Rat ornithine decarboxylase pseudogene.
ACCESSION	X13417
NID	956788
VERSION	X13417.1 GI:56788
KEYWORDS	ornithine decarboxylase; pseudogene.
SOURCE	Norway rat.
ORGANISM	Rattus norvegicus
REFERENCE	Eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS	Rodenilla; Schiurognath; Muridae; Murinae; Rattus.
TITLE	1 (bases 1 to 3093)
JOURNAL	Oka, T.
REFERENCE	Submitted (03-NOV-1988) to the EMBL/GenBank/DBJ databases
AUTHORS	2 (bases 1 to 3093)
TITLE	Kanamoto, R., Yoshimura, M., Hayashi, S. and Oka, T.
JOURNAL	Nucleotide sequence of a pseudogene for rat ornithine decarboxylase
MEDLINE	Nucleic Acids Res. 17 (1), 463 (1989)
COMMENT	89096420
FEATURES	Data kindly reviewed (20-Apr-1989) by Oka T.
SOURCE	Location/Qualifiers
	1..3093

Query Match	23.7%	Score 34.4	DB 12	Length 3093
Best Local Similarity	57.4%	Pred. NC. 0.2		
Matches 62	Conservative	0	Mismatches 46	Indels 0
				Gaps 0

QY 1 CTAATTTCACATTTACCAAGTGGACACCTGAGTACAGCGAGACCTGGAAATCCCAAGATGAGAGA 60  
Db 1488 CTCCTTCACACCTTCACACTGGGGAATGATGTACACTGCAGAGTGTAAAGTAGAGAGACA 1429

RESULT	4	
MINCO1/C		
LOCUS	491 bp	DNA
DEFINITION	<i>Mimospora crassa</i> ATPase proteolipid like gene and unidentified reading frame on complementary strand.	21-OCT-1996
ACCESSION	V00667	J01428
NID	92938	
VERSION	V00667.1	GI:2938

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:16 ; Search time 2825.05 seconds

(without alignments)  
58.651 Million cell updates/sec

Title: US-08-956-518a-97

Perfect score: 84  
Sequence: 1 TTAAACACAGATATGTAAC.....GCATTTACAGTAGATCAT 84

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

abase :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: em\_est20: \*  
21: em\_est21: \*  
22: em\_est22: \*  
23: em\_est23: \*  
24: em\_est24: \*  
25: em\_est25: \*  
26: em\_est26: \*  
27: em\_est27: \*  
28: em\_est28: \*  
29: em\_est29: \*  
30: em\_est30: \*  
31: em\_est31: \*  
32: em\_est32: \*  
33: em\_est33: \*  
34: em\_est34: \*  
35: em\_est35: \*  
36: em\_est36: \*  
37: em\_est37: \*  
38: em\_est38: \*  
39: em\_est39: \*  
40: em\_est40: \*  
41: em\_est41: \*  
42: em\_est42: \*  
43: em\_est43: \*  
44: em\_est44: \*  
45: em\_est45: \*  
46: em\_est46: \*  
47: em\_est47: \*  
48: em\_est48: \*  
49: em\_est49: \*  
50: em\_est50: \*  
51: em\_est51: \*  
52: em\_est52: \*  
53: em\_est53: \*

54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45.2	53.8	501	49	A1655990
2	38	45.2	500	26	W52861
3	31.6	37.6	487	39	AA861176
4	31.2	37.1	457	44	AI286133
5	28.8	34.3	549	34	AA475853
6	28.2	33.6	420	49	AT631518
7	28	33.3	567	49	AU056388
8	27.8	33.1	362	21	D47311
9	27.6	32.9	395	23	H47049
10	27.6	32.9	353	23	R93050
11	27.6	32.9	202	32	AA360372
12	27.6	32.9	430	32	AA398114
13	27.6	32.9	493	35	AA553441
14	27.4	32.6	496	35	AA572984
15	27.2	32.4	567	48	AU051905
16	27	32.1	557	49	AI622332
17	27	32.1	696	49	AU057033
18	26.6	31.7	563	49	AT651377
19	26.4	31.4	322	20	T40435
20	26.4	31.4	379	21	R09085
21	26.4	31.4	366	21	T80555
22	26.4	31.4	506	22	R66628
23	26.4	31.4	449	24	H98196
24	26.4	31.4	449	24	H99885
25	26.4	31.4	451	25	N55011
26	26.4	31.4	441	25	W12683
27	26.4	31.4	582	26	W54506
28	26.4	31.4	495	26	W62896
29	26.4	31.4	468	26	W77761
30	26.4	31.4	580	27	AA003956
31	26.4	31.4	512	27	AA010721
32	26.4	31.4	582	27	AA016534
33	26.4	31.4	619	27	AA030762
34	26.4	31.4	573	28	AA075433
35	26.4	31.4	661	28	AA124475
36	26.4	31.4	496	31	AA310074
37	26.4	31.4	464	31	AA314005
38	26.4	31.4	280	32	AA334911
39	26.4	31.4	344	32	AA340148
40	26.4	31.4	310	32	AA344762
41	26.4	31.4	199	32	AA373126
42	26.4	31.4	660	33	AA394081
43	26.4	31.4	556	33	AA405892
44	26.4	31.4	370	33	AA442497
45	26.4	31.4	826	50	AU067186

#### ALIGNMENTS

RESULT 1  
LOCUS A1655990/c  
DEFINITION tt42c03.x1 NCI\_CGAP\_G66 Homo sapiens cDNA clone IMAGE:2243428 3',  
ACCESSION A1655990  
NID 94739969  
VERSION A1655990.1 GI:4739969

EST 04-MAY-1999

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE 1 (bases 1 to 501)  
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jun 5, 1998 this sequence version replaced gi:3188057.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: [Robert.Strausberg@nih.gov](mailto:Robert.Strausberg@nih.gov)  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.llnl.gov/bdrrp/image/image.html](http://www.bio.llnl.gov/bdrrp/image/image.html)

Seq primer: -40up from G1bco  
High quality sequence stop: 458.  
Location/Qualifiers  
1..501  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3243428"  
/clone\_1lb="NCI-CCAP\_GC6"  
/tissue\_type="Pooled germ cell tumors"  
/lab\_host="DH10B"  
/note="Vector: pRTT3D-Pac (Pharmacia) with a modified polylinker. Plasmid DNA from the normalized library NCI-CCAP\_GC6 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 135 a 107 c 115 g 144 t  
ORIGIN

Query Match 53.8%; Score 45.2; DB 49; Length 501;  
Best Local Similarity 90.9%; Pred. No. 0.00066;  
Matches 60; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

QY 9 AGATATGAAACACCATCGTTAAATTGATGCAAAATATTGATCATTACAGCAT 68  
|||||  
Db 501 AAGATATGAAACAGCACCA---GTTAAATTGATGCAAAATATTGATCATTACAGCAT 445

QY 69 TTTTCAG 74  
|||||  
Db 444 ATTTCAG 439

RESULT 2  
W52861 560 bp mRNA EST 10-OCT-1996  
LOCUS zc03e02.r1 Soares-parathyroid\_tumor\_NbHPA Homo sapiens cDNA clone  
IMAGE:321242.5, similar to gb:X70297 NEURONAL ACETYLCHOLINE  
RECEPTOR PROTEIN, ALPHA-7 CHAIN (HUMAN), mRNA sequence.  
ACCESSION W52861  
NID G1350351  
VERSION W52861.1 GI:1350351  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE 1 (bases 1 to 560)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Matra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.  
The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT On Jan 24, 1995 this sequence version replaced gi:634363.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert length: 1753 Std Error: 0.00  
Seq primer: mob.REGA+ET  
High quality sequence stop: 408.  
Location/Qualifiers  
1..360  
/organism="Homo sapiens"  
/db\_xref="GDB:1258900"  
/db\_xref="taxon:9606"  
/clone="IMAGE:321242"  
/clone\_1lb="Soares-parathyroid\_tumor\_NbHPA"  
/tissue\_type="parathyroid tumor"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: parathyroid gland; Vector: pRTT3D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer  
15'-  
TGTTCACATCTGAGAGTGGAGCGCCGACCAATTTTCTTTTCTTTTCTTTT  
T-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRTT3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH."

BASE COUNT 128 a 137 c 137 g 152 t 6 others  
ORIGIN

Query Match 45.2%; Score 38; DB 26; Length 560;  
Best Local Similarity 89.1%; Pred. No. 0.076;  
Matches 41; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 36 AATTGATGCAAAATATTGATCATTACAGCATTTTCAGTAGGAT 81  
|||||  
Db 1 AATTGATGCAAAATATTGATCATTACAGCATTTTCAGTTCCAT 46

RESULT 3  
AA861176 487 bp mRNA EST 04-JAN-1999  
LOCUS AK36409.s1 Soares\_testis\_NHT Homo sapiens cDNA clone IMAGE:1408049  
DEFINITION 3', mRNA sequence.  
ACCESSION AA861176  
NID G2953316  
VERSION AA861176.1 GI:2953316  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE 1 (bases 1 to 487)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jan 9, 1998 this sequence version replaced gi:3937098.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
[www-bio.lnl.gov/db/ftp/image/image.html](http://www-bio.lnl.gov/db/ftp/image/image.html)

Insert Length: 758 Std Error: 0.00  
 Seq primer: -40m13 fwd. ET from Amersham  
 High quality sequence stop: 458.  
 Location/Qualifiers

## FEATURES

## source

1. 487  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:140804.9"  
 /clone\_id="Soares\_testis\_NHT"  
 /sex="male"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from Clontech  
 Laboratories, Inc., and primed with a Not I - oligo(dT)  
 primer [5].  
 TGTACCAATCGATGAGGAGGCGGCCCAATTTTCTTTTCTTTT 3']  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization to Cot5, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT  
 ORIGIN

131 a 95 c 102 g 159 t

Query Match 37.6%; Score 31.6; DB 39; Length 487;  
 Best Local Similarity 89.5%; Pred. No. 5.2;  
 Matches 34; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

44 GCAAAATATGTCATCCAGCATTTTTCAGTAGAT 81  
 |||||  
 Db 487 GCAAAATATGTCATCCAGCATTTTTCAGTCCAT 450

RESULT 4  
 LOCUS A1286133  
 DEFINITION q101c05.x1 Soares.NFL.T.GBC.S1 Homo sapiens CDNA clone  
 IMAGE:1855208 3' similar to gb:U74525 UBIQUITIN-CONJUGATING ENZYME  
 E2-17 KD (HUMAN); mRNA sequence.  
 ACCESSION A1286133  
 NID 93924366  
 VERSION A1286133.1 GI:3924366  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 457)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)

COMMENT On Jan 17, 1998 this sequence version replaced gi:2044002.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([infoimage.lnl.gov](http://infoimage.lnl.gov)) for further information.  
 Insert Length: 787 Std Error: 0.00  
 Seq primer: -40UP from Gldco.  
 High quality sequence stop: 370.  
 Location/Qualifiers

## FEATURES

## source

1. 457  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1855208"  
 /clone\_id="Soares\_NFL.T.GBC.S1"  
 /lab\_host="DH10B"  
 /note="Organ: pooled. Vector: pT73D-Pac (Pharmacia) with  
 a modified polylinker. Site\_1: Not I; Site\_2: Eco RI;  
 Equal amounts of plasmid DNA from three normalized  
 libraries (fetal lung NBHL19W, testis NHT, and B-cell  
 NCI-CGAP\_GCB1) were mixed, and ss circles were made in  
 vitro. Following HAP purification, this DNA was used as  
 tracer in a subtractive hybridization reaction. The driver  
 was PCR-amplified cDNAs from pools of 5,000 clones made  
 from the same 3 libraries. The pools consisted of  
 I.M.A.G.E. clones 297480-302087, 682632-687239,  
 726408-728711, and 729096-731399. Subtraction by Bento  
 Soares and M. Fatima Bonaldo."

BASE COUNT  
 ORIGIN

128 a 89 c 94 g 144 t 2 others

Query Match 37.1%; Score 31.2; DB 44; Length 457;  
 Best Local Similarity 63.2%; Pred. No. 6.8;  
 Matches 48; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 9 AGATATGAAACACCAATCGTTAAATTTGATGCACCAAAATATGTCATCCAGCAT 68  
 |||  
 Db 399 AGATATGAAATACCAATCCAGTTAGATTGTCTTAAGATGTTCCATCCAAATGCT 340  
 QY 69 TTTCAGGTAGATCAT 84  
 |||  
 Db 339 ATCCAGATGCTAGTAT 324

RESULT 5  
 LOCUS AA475853  
 DEFINITION vhl1g04.r1 Soares mouse mammary gland NBMMG Mus musculus CDNA clone  
 IMAGE:875286 5', mRNA sequence.  
 ACCESSION AA475853  
 NID 92203704  
 VERSION AA475853.1 GI:2203704  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 549)  
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisler,S., Kucada,T., Lacy,M., Le,M., Martin,D., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT On Nov 29, 1993 this sequence version replaced gi:430168.

Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108



RESULT  
H47049

Query Match	32.98;	Score 27.6;	DB 23;	Length. 395;
Best Local Similarity	61.88;	Pred. No. 75;		
Matches 42;	Conservative 0;	Mismatches 26;	Indels 0;	Gaps 0;



Query Match	32.9%	Score 27.6;	DB 23;	Length 353;
Best local Similarity	61.8%	Pred. No. 75;		
Matches 42;	Conservative 0;	Mismatches 26;	Indels 0;	Gaps 0

Query Match	32.98;	Score 27.6;	DB 32;	Length 202
Best Local Similarity	61.88;	Pred. No. 80;		

[illegible]

**CONTACT:** Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
**TEL:** 314 286 1800  
**FAX:** 314 286 1810  
**EMAIL:** estewartson.wustl.edu  
 This clone is available royalty-free through INLNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
**SEQ PRIMER:** -4m13 fwd. ET from Aneersham  
 High quality sequence stop: 142.  
**LOCATION/QUALIFIERS**  
 1. .430

```

/organism="Homo sapiens"
/db_xref="GDB:5923466"
/db_xref="taxon:9606"
/map="13"
/clone="IMAGE:726554"
/clone_id="Soares_Testis_NHT"
/sex="male"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5]
TGTACCAATCTGACATGGAGCGGCCGCCCAATTTTTTTTTTTT 3'.
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot's and was
constructed by Bento Soares and M. Fatima BonaIdo. "

```

Query Match	32.9%;	Score 27.6;	DB 33;	Length 430;
Best Local Similarity	60.8%;	Pred. No. 74;		
Matches 45;	Conservative	0;	Mismatches 29;	Indels 0;
				Gaps 0;

Oy 11 AATATGAAACAAACCCATCGGTAAATTGATGCAAAAAATATGCATCTACAGCATTT 70  
 Db 412 AATACCAATTAACACACCTCACTAGATTGTCTCTAAGATGTTCCATCCAAATGTCTAT 35  
 Oy 71 TCAAGTAGATCAT 84  
 Db 352 GCAGATGTGATAT 339

RESULT	13
AA553441/c	
LOCUS	AA553441
DEFINITION	kR7d07.s1 NCI-CGAP_Homo sapiens CDNA clone IMAGE:10195333
SOURCE	Similar to gb:M5663 INTERFERON-INDUCED, DOUBLE-STRAINED RNA-ACTIVATED PROTEIN KINASE (HUMAN); mRNA sequence.
ACCESSION	AA553441
NID	g2323980
VERSION	AA553441.1 GI:2323980
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Euarcheria; Primates; Catarrhini; Homidae; Homo. 1 (bases 1 to 493) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)
JOURNAL COMMENT	On Apr 14, 1993 this sequence version replaced gi:692969.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.  
Zimmer-Buck, M.D., Ph.D.  
CDNA Library Preparation: Stratagene, Inc., David B. Krizman,  
Ph.D.  
CDNA Library Arraying: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
clone distribution: NCI-CCGAP clone distribution Information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/dbfp/image/image.html](http://www-bio.llnl.gov/dbfp/image/image.html)

```

FEATURES
source
1..493
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="MMU16C3-C4 region; MMU16C3-C4 region"
/clone_image="1019533"
/clone_id="NCI_CGAP_Sch1"
/tissue_type="Schwannoma tumor"
/lab_host="SOLR (Kanamycin resistant)"
/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dt. Two
pooled bulk schwannoma tumors. 5 adaptor sequence: 5'
GAAATTCGACGAG 3' adaptor sequence: 5'
CTGACGATTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."
BASE COUNT
152 a 97 c 72 g 172 t
ORIGIN

```

Query Match	32.9%;	Score 27.6;	DB 35;	Length 493;
Best Local Similarity	60.8%;	Pred. No. 73;		
Matches 45; Conservative	0;	Mismatches 29;	Indels 0;	Gaps 0;

QY       7 AAGAGTAATGAAACACACACATCGGTAAATTGGATGCAAAAATTTGCATCTPCCAGC 66  
          | | | | | | | | | | | | | | | |  
Db     207 AAAGCAAAAAAAGTAAACATTTGGTGATTTATGTAGATAATAAACATATGCCAAC 148  
          | | | | | | | | | | | | | | | |  
QY       67 ATTTCAGGTAGA 80  
          | | | | | | | |



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:03:54 ; Search time 1962.17 Seconds

(without alignments)  
108.594 Million cell updates/sec

Title: US-08-956-518a-98

Sequence: 1 TTTATCTGATGTCGCAATTC.....ACTGCATATGTAAGTACA 67

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Abase :

GenEmb1: \*  
1: gb\_da1: \*  
2: gb\_da2: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_pl1: \*  
8: gb\_pl2: \*  
9: gb\_pl3: \*  
10: gb\_pr2: \*  
11: gb\_pr3: \*  
12: gb\_ro: \*  
13: gb\_st: \*  
14: gb\_sts: \*  
15: gb\_sy: \*  
16: gb\_un: \*  
17: gb\_vl: \*  
18: em\_fun: \*  
19: em\_hcg: \*  
20: em\_hum1: \*  
21: em\_hum2: \*  
22: em\_in: \*  
23: em\_om: \*  
24: em\_or: \*  
25: em\_ov: \*  
26: em\_pat: \*  
27: em\_ph: \*  
28: em\_pl: \*  
29: em\_ro: \*  
30: em\_sts: \*  
31: em\_sy: \*  
32: em\_un: \*  
33: em\_vl: \*  
34: gb\_hcg1: \*  
35: gb\_hcg2: \*  
36: gb\_in1: \*  
37: gb\_in2: \*  
38: em\_da1: \*  
39: em\_da2: \*  
40: em\_hum3: \*  
41: em\_hum4: \*  
42: gb\_pr4: \*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	57	85.1	1836 11	AF037646 Homo sapi

2	49.8	74.3	641	11	AF029838	AF029838 Homo sapi
3	49.8	74.3	1712	11	AF036903	AF036903 Homo sapi
4	47.4	70.7	689	11	AF029839	AF029839 Homo sapi
5	29.2	43.6	163403	37	AC007765	AC007765 Drosophila
6	28.8	43.0	167525	11	AC006531	AC006531 Homo sapi
7	26.8	40.0	7434	17	HPV9	X74464 Human papill
8	26.8	40.0	169891	35	AC007037	AC007037 Homo sapi
9	26.6	39.7	1084	12	MUSMRNA	L36962 Mus musculu
10	25.8	38.5	161336	11	AC004067	AC004067 Homo sapi
11	25.8	38.5	2638	12	AB015894	L03367 Human Oligo
12	25.8	38.5	2638	12	AB015894	AB015894 Mus muscu
13	25.8	38.5	2638	12	AB015894	AB015894 Mus muscu
14	25.8	38.5	2638	12	AB015894	AB015894 Mus muscu
15	25.6	38.2	141313	11	AC004943	AC004943 Homo sapi
16	25.6	38.2	54705	34	AC006094	AC006094 Homo sapi
17	25.4	37.9	8393	1	U39732	U39732 Mycoplasma
18	25.4	37.9	9616	2	U39727	U39727 Mycoplasma
19	25.4	37.9	100625	10	HS1189K21	AL030997 Human DNA
20	25.4	37.9	89097	42	AC005518	AC005518 Homo sapi
21	25.4	37.9	28945	42	AF048729	AF048729 Homo sapi
22	25.2	37.6	74842	7	AB016877	AB016877 Arabidops
23	25.2	37.6	4030	7	POTST4C11	M62755 Potato 4-co
24	25.2	37.6	140094	10	HS2013	AL035423 Human DNA
25	25.2	37.6	85530	34	HS469D22	AL031284 Homo sapi
26	25.2	37.6	32424	11	AC004037	AC004037 Homo sapi
27	24.8	37.0	183085	11	AC005815	AC005815, complet
28	24.6	36.7	3882	7	D50797	D50797 Schizosacch
29	24.6	36.7	1361	7	HASF21	Y09057 H. annuus MR
30	24.6	36.7	42037	7	SPAC23D3	Z64354 S. pombe chr
31	24.6	36.7	81467	8	ATAC004683	AC004683 Arabidops
32	24.6	36.7	1522	36	TBRCTJ2G	L42549 Trypanosoma
33	24.4	36.4	68098	7	AB020752	AB020752 Arabidops
34	24.4	36.4	1975	12	MUSHRNPK	L29769 House mouse hnrnp
35	24.4	36.4	126983	34	AC005283	AC005283 Homo sapi
36	24.4	36.4	155313	8	AC006453	AC006453 Homo sapi
37	24.4	36.1	6513	8	SIU55278	U55278 Solanum lyc
38	24.2	36.1	100000	9	AP000147	AP000147 Homo sapi
39	24.2	36.1	89086	9	AP000233	AP000233 Homo sapi
40	24.2	36.1	121290	34	CEH27C14	Z98852 Caenorhabdi
41	24.2	36.1	300172	35	AC005308	AC005308 Plasmodiu
42	24.2	36.1	25282	36	CEC30F2	Z70681 Caenorhabdi
43	24.2	36.1	39353	36	CEP11A1	Z50857 Caenorhabdi
44	24.2	36.1	119996	42	AC006155	AC006155 Homo sapi
45	24.2	36.1	119996	42	AC006155	AC006155 Homo sapi

## ALIGNMENTS

RESULT 1  
AF037646  
LOCUS  
DEFINITION Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor precursor RNA, partial sequence.  
ACCESSION AF037646  
NID 93757808  
VERSION AF037646.1 GI:3757808  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J., Moore, T., Jacobs, S., Meriwether, J., Choi, M. J., Kim, E. J., Walton, K., Bulting, R., Davis, A., Breese, C., Freedman, R. and Leonard, S.  
TITLE Genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene (CHRNA7)  
JOURNAL JOURNAL MEDLINE  
99000837  
2 (bases 1 to 1836)  
REFERENCE  
AUTHORS Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R., Breese, C., Davis, A., Hopkins, J. and Freedman, R.

```

TITLE      Direct Submission
JOURNAL    Submitted (03-DEC-1997) Psychiatry, University of Colorado Health
            Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
FEATURES   source
            1. 1836
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /db_xref="dbEST:W03952"
                /db_xref="dbEST:N73891"
                /chromosome="15"
                /clone="297134"
                /clone_1lb="Scars fetal liver spleen library, Research
                Genetics/IMAGE Consortium, LBN"
                /note="unprocessed mRNA with intron"
            <1. 1836
                /note="alpha-7 neuronal nicotinic acetylcholine receptor
                precursor; intron included in 5' region"

BASE COUNT      383 a      503 c      469 g      481 t
ORIGIN

misc.feature

Query Match      85.1%; Score 57; DB 11; Length 1836;
Best Local Similarity 100.0%; Pred. No. 6e-10;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1      TTTATTCTAGTTCACATTGCTTATTCACAGATTGTGGAGTACGTGCNAACGTGGATAT 57
Db      160      TTTATTCTAGTTCACATTGCTTATTCACAGATTGTGGAGTACGTGCNAACGTGGATAT 216

RESULT 2
LOCUS      AF029838      641 bp      mRNA      PRI      16-DEC-1998
DEFINITION Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.
ACCESSION  AF029838
VERSION     93757793
SOURCE      AF029838.1 GI:3757793
ORGANISM   human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 641)
AUTHORS     Gault,J., Robinson,M., Berger,R., Drebing,C., Logel,J., Hopkins,J.,
            Moore,T., Jacobs,S., Meliether,J., Choi,M.J., Kim,E.J., Walton,K.,
            Bulling,K., Davis,A., Breese,C., Freedman,R. and Leonard,S.
            Genomic organization and partial duplication of the human alpha7
            neuronal nicotinic acetylcholine receptor gene
            Genomics 52 (2), 173-185 (1998)
JOURNAL     2 (bases 1 to 641)
REFERENCE   Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R.,
AUTHORS     Breese,C., Davis,A., Hopkins,J. and Freedman,R.
            Direct Submission
            Submitted (15-OCT-1997) Psychiatry, University of Colorado Health
            Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
FEATURES   Location/Qualifiers
            1. 641
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="15"
                /map="15q14"
                /clone="CHRNA7-DR1"
                /note="RACE product A/B/C/D"
            <1. >641
                /note="alpha 7 neuronal nicotinic receptor mRNA; used in
                alternative splicing"

BASE COUNT      151 a      192 c      163 g      135 t
ORIGIN

Query Match      74.3%; Score 49.8; DB 11; Length 641;
Best Local Similarity 96.2%; Pred No. 1.9e-07;

```

Matches	51: Conservative	0: Mismatches	2: Indels	0: Gaps	0: Indels
Qy	5	TTTCAGTTCGCAATTCGTAATCCAGCAATTTGTGATAGCTGCAAACTGCGATAT	57		
Db	482	TTTCAGTTCGCAATTCGTAATCCAGCAATTTGTGATAGCTGCAAACTGCGATAT	534		
RESULT	3				
AF036903	1712 bp	mRNA	PRI	16-OCT-1998	
LOCUS	Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor				
DEFINITION	mRNA, alternatively spliced, partial sequence.				
ACCESSION	AF036903				
NID	93757807				
VERSION	AF036903.1	GI:3757807			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Primates; Catarrhini; Homiidae; Homo.				
TITLE	1 (bases 1 to 1712)				
JOURNAL	Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J., Moore, T., Jacobs, S., Meriwether, J., Choi, M., Kim, E., Walton, K., Bulding, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.				
MEDLINE	Genomic organization and partial duplication of the human alpha7				
REFERENCE	neomonic nicotinic acetylcholine receptor gene (CHRNA7)				
AUTHORS	Genomics 52 (2), 173-185 (1998)				
TITLE	2 (bases 1 to 1712)				
JOURNAL	Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R., Breese, C., Davis, A., Hopkins, J. and Freedman, R.				
MEDLINE	Submitted (03-DEC-1997) Psychiatry, University of Colorado Health				
REFERENCE	Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA				
FEATURES	location/qualifiers				
SOURCE	1..1712				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/db_xref="dbest:W52861"				
	/db_xref="dbest:AA037389"				
	/chromosome="15"				
	/clone="321242"				
	/clone_11b="Scores papathryoid tumor NBHPA library,				
	Research Genetics/IMAGE Consortium, LNL"				
	<1..1712				
	/note="alpha-7 neuronal nicotinic acetylcholine receptor;				
	alternatively spliced"				
BASE COUNT	356 a 489 c 457 g 410 t				
ORIGIN					
misc_feature					
Query Match	74.3%; Score 49.8; DB 11; Length 1712;				
Best Local	Similarity 96.2%; Pred. No. 1.8e-07;				
Matches	51: Conservative 0; Mismatches 2; Indels 0; Gaps 0				
Qy	5	TTTCAGTTCGCAATTCGTAATCCAGCAATTTGTGATAGCTGCAAACTGCGATAT	57		
Db	42	TTTCAGTTCGCAATTCGTAATCCAGCAATTTGTGATAGCTGCAAACTGCGATAT	94		
RESULT	4				
AF029839	689 bp	mRNA	PRI	16-DEC-1998	
LOCUS	Homo sapiens alpha 7 neuronal nicotinic acetylcholine receptor				
DEFINITION	mRNA sequence.				
ACCESSION	AF029839				
NID	93757794				
VERSION	AF029839.1	GI:3757794			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Primates; Catarrhini; Homiidae; Homo.				
TITLE	1 (bases 1 to 689)				

**AUTHORS**  
Gault, J., Robinson, M., Berger, R., Drebling, C., Logel, J., Hopkins, J., Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K., Buling, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.

**TITLE**  
Genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene

**JOURNAL**  
Genomics 52 (2), 173-185 (1998)

**MEDLINE**  
99000837

**REFERENCE**  
2 (bases 1 to 689)

**AUTHORS**  
Leonard, S., Gault, J., Logel, J., Drebling, C., Robinson, M., Berger, R., Breese, C., Davis, A., Hopkins, J. and Freedman, R.

**TITLE**  
Direct Submission

**JOURNAL**  
Submitted (15-OC-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA

**FEATURES**  
Location/Qualifiers  
1..689  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15q14"  
/clone="CHRNA7-DR2"  
/note="RACE product A/C/D"  
1..>689  
/note="alpha 7 neuronal nicotinic receptor mRNA; used in alternative splicing"

**BASE COUNT**  
142 a 210 c 196 g 141 t

**ORIGIN**

**Query Match**  
Best Local Similarity 70.7%; Score 47.4; DB 11; Length 689;  
Matches 48; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**QY**  
9 AGTCCATTGCTAATCCAGCATTTGGATAGCTGCACAACTGCATAT 57  
1 |||||  
Db 534 AATTCCATTGCTAATCCAGCATTTGGATAGCTGCACAACTGCATAT 582

**RESULT**  
5 AC007765 163403 bp DNA INV 09-JUN-1999  
LOCUS Drosophila melanogaster, chromosome 2L, region 23C1-23C5, P1 clones  
DEFINITION DS02190 and DS00906, complete sequence.  
AC007765 AC004375 AC003632 AC003633 AC003634 AC003635  
AC004134 AC002028 AC002976 AC002029 AC002977 AC002030 AC002031  
AC002032 AC002033 AC002034 AC002035  
5030435  
NID AC007765.1 GI:5030435  
SOURCE HTG.  
WORDS  
ORGANISM fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 163403)  
Celinker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,  
Butenkov, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,  
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,  
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,  
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snit, E.,  
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
Rubin, G.M.

**TITLE**  
Sequencing of Drosophila chromosome 2L, region 23C1-23C5  
Unpublished (1998)

**JOURNAL**  
2 (bases 1 to 163403)

**REFERENCE**  
AUTHORS Celinker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,  
Butenkov, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,  
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,  
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,  
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snit, E.,  
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
Rubin, G.M.

**TITLE**  
Direct Submission

**JOURNAL**  
Submitted (09-JUN-1999) Berkeley Drosophila Genome Project, MS  
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,  
Berkeley, CA 94720, US  
On Jun 9, 1999, this sequence version replaced gi:3293203  
gi:3097820.

**COMMENT**  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdgs@fruitfly.berkeley.edu](mailto:bdgs@fruitfly.berkeley.edu).  
P1 library locations: 23-78, 10-42.

**FEATURES**  
Location/Qualifiers  
1..163403  
/organism="Drosophila melanogaster"  
/strain="y; cn bw<sup>sp</sup>"  
/db\_xref="taxon:7227"  
/chromosome="2L"  
/map="23C1-23C5"  
/clone="11b-P1 library, partial Sau3A in Padi05acB11"  
/clone="P1s DS02190 (D82) and DS00906 (D99)"  
/note=" "

**BASE COUNT**  
45206 a 35896 c 36551 g 45750 t

**ORIGIN**

**Query Match**  
Best Local Similarity 43.6%; Score 29.2; DB 37; Length 163403;  
Matches 43; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

**QY**  
2 TTAATCTAGTCCATTGCTAATCCAGCATTTGGATAGCTGCACAACTGCATATGAA 61  
1 |||||  
Db 108123 TTAATCTAGTCCATTGCTAATCCAGCATTTGGATAGCTGCACAACTGCATATGAA 108182

**RESULT**  
6 AC006531 167525 bp DNA PRI 06-FEB-1999  
LOCUS Homo sapiens chromosome 16 clone 113K5, complete sequence.  
DEFINITION AC006531  
AC006531  
NID g4235137  
VERSION AC006531.1 GI:4235137  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 167525)  
Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,  
Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,  
Bryant, J., Tesmer, J., Melnick, L., Longmire, J., White, S., Tatum, O.,  
Campbell, C., Fawcett, J., Maltbie, M., Bussod, M., Sutherland, R.,  
Han, C. and Deaven, L.

**TITLE**  
Sequencing of Human Chromosome 16p13.3  
Unpublished

**JOURNAL**  
2 (bases 1 to 167525)

**REFERENCE**  
AUTHORS Rucke, D.O.

**TITLE**  
Large Scale Sequence Analysis and Annotation with the Sequence  
Comparison Analysis (SCAN) System

**JOURNAL**  
Unpublished

**REFERENCE**  
3 (bases 1 to 167525)

**AUTHORS**  
Jones, M., Mundt, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,  
Bruce, D., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,  
Bryant, J., Tesmer, J., Melnick, L., Longmire, J., White, S., Tatum, O.,  
Campbell, C., Fawcett, J., Maltbie, M., Bussod, M., Sutherland, R.,  
Han, C. and Deaven, L.

Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,  
McNulty,K., Han,C. and Deaven,L.  
Direct Submission  
Submitted (06-FEB-1999) Center for Human Genome Studies, DOE Joint  
Genome Institute, Los Alamos National Laboratory, MS M888, Los  
Alamos, NM 87545, USA

## FEATURES

## SOURCE

```

1. 167525
Location/Qualifiers
misc_feature
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="113K5"
    /chromosome="16"
    2396..2466
repeat_region
    /rpt_family="MIR"
    complement(join(3760..10746,14646..14885,44104..44260,
68059..68223,75207..75450,79902..80033,86444..86599,
86723..86902,95554..95760))
    /note="NM2A receptor modulatory subunit"
    12465..12543
repeat_region
    /rpt_family="Alu"
    complement(13201..13294)
repeat_region
    /rpt_family="L1"
    complement(13836..14146)
misc_feature
    /rpt_family="Alu"
    complement(14600..14851)
    /note="GRAIL 2 excellent exon, frame 2"
    complement(16599..16853)
repeat_region
    /rpt_family="L1"
    complement(18201..18458)
repeat_region
    /rpt_family="Alu"
    23577..23860
repeat_region
    /rpt_family="Alu"
    complement(24365..25041)
repeat_region
    /rpt_family="MER"
    26799..26960
repeat_region
    /rpt_family="MER"
    complement(27162..27453)
repeat_region
    /rpt_family="Alu"
    28101..28204
repeat_region
    /rpt_family="MER2"
    29619..29862
repeat_region
    /rpt_family="L1"
    29943..30269
repeat_region
    /rpt_family="Alu"
    30282..31394
repeat_region
    /rpt_family="L1"
    32685..32828
repeat_region
    /rpt_family="MIR"
    33206..33361
repeat_region
    /rpt_family="MER4"
    complement(34052..34381)
repeat_region
    /rpt_family="Alu"
    complement(34520..34772)
repeat_region
    /rpt_family="Alu"
    35856..36026
    /note="GRAIL 2 excellent exon, frame 1"
    complement(38026..38318)
repeat_region
    /rpt_family="Alu"
    38718..38821
repeat_region
    /rpt_family="MER5"
    complement(40648..40860)
    /note="GRAIL 2 excellent exon, frame 1"
    complement(40863..41382)
repeat_region
    /rpt_family="MER41"
    complement(43625..43725)
    /rpt_family="MER20"
    complement(44039..44226)
    /note="GRAIL 2 excellent exon, frame 2"
    45220..45336
repeat_region
    /rpt_family="MLT"
    45337..45419
repeat_region
    /rpt_family="MER"
    complement(45749..46003)

```

---

```

    /rpt_family="Alu"
    46499..46629
    /note="GRAIL 2 excellent exon, frame 0"
    47808..48246
repeat_region
    /rpt_family="L1"
    48509..48629
repeat_region
    /rpt_family="MER3"
    48919..49218
repeat_region
    /rpt_family="Alu"
    complement(49485..49635)
repeat_region
    /rpt_family="MIR"
    51405..51488
repeat_region
    /rpt_family="MIR"
    52326..52469
    /note="GRAIL 2 excellent exon, frame 1"
    52626..52963
repeat_region
    /rpt_family="THE1"
    complement(53966..54143)
    /rpt_family="MIR"
    complement(54381..54442)
    54933..55231
repeat_region
    /rpt_family="Alu"
    55579..55875
repeat_region
    /rpt_family="Alu"
    56250..56595
repeat_region
    /rpt_family="MLT1"
    57272..57582
repeat_region
    /rpt_family="Alu"
    58153..58480
repeat_region
    /rpt_family="Alu"
    60534..60842
repeat_region
    /rpt_family="Alu"
    61134..61527
repeat_region
    /rpt_family="L1"
    61633..61910
repeat_region
    /rpt_family="L1"
    61958..62058
repeat_region
    /rpt_family="L1"
    62540..62679
repeat_region
    /rpt_family="MER5"
    64727..64830
repeat_region
    /rpt_family="L1"
    65034..65336
repeat_region
    /rpt_family="Alu"
    65464..65742
repeat_region
    /rpt_family="Alu"
    complement(68027..68187)
    /note="GRAIL 2 excellent exon, frame 1"
    complement(69367..69526)
repeat_region
    /rpt_family="MIR"
    complement(75186..75415)
    /note="GRAIL 2 excellent exon, frame 2"
    complement(75606..75636)
    /note="GRAIL 2 excellent exon, frame 2"
    complement(77284..77625)
repeat_region
    /rpt_family="Alu"
    complement(77735..77896)
repeat_region
    /rpt_family="Alu"
    complement(77916..78047)
repeat_region
    /rpt_family="Alu"
    complement(79531..79812)
    /rpt_family="Alu"
    complement(79868..79993)
    /note="GRAIL 2 excellent exon, frame 2"
    complement(80168..80463)
repeat_region
    /rpt_family="Alu"
    84026..84320
repeat_region
    /rpt_family="L1"
    84674..84820
repeat_region
    /rpt_family="L1"
    85690..85917
    /rpt_family="Alu"

```

```

repeat_region      /rpt_family="RHE1"      complement(865933, .86272)
misc_feature        /note="GRAIL 2 excellent exon, frame 0"      complement(86610, .86544)
misc_feature        /note="GRAIL 2 excellent exon, frame 0"      complement(86699, .86867)
repeat_region      /note="GRAIL 2 excellent exon, frame 0"      complement(86699, .86867)
repeat_region      /rpt_family="Alu"      87763, .88083
repeat_region      /rpt_family="Alu"      88459, .88774
repeat_region      /rpt_family="Alu"      89136, .89292
repeat_region      /rpt_family="Alu"      93079, .93370
repeat_region      /rpt_family="Alu"      93865, .94055
repeat_region      /rpt_family="MER7"      complement(93994, .94311)
misc_feature        /rpt_family="Alu"      complement(95519, .95774)
repeat_region      /note="GRAIL 2 excellent exon, frame 2"      96104, .98020
repeat_region      /rpt_family="MIR"      complement(98641, .98718)
repeat_region      /rpt_family="L1"      99052, .99153
repeat_region      /rpt_family="MLT1"      99315, .99414

Query Match      43.0%; Score 28.8; DB 11; Length 167525;
Best Local Similarity 65.6%; Pred. NO. 1.9; Mismatches 22; Indels 0; Gaps 0;
Matches 42; Conservative 0;

QY      4      ATTCAGTCCAAATGCTAATCCAGCAATTTGGATAGCTGCAAACTGCGATATGTAAT 63
Db 146832      ATTCAGAGTGGCAATTTATTTCAGAAATTTCTGATATTCACATACTGGAATATAGAAAT 146891
QY      64      AACCA 67
Db 146892      CATA 146895

RESULT      7
HPV9      7424 bp      DNA      VRL      10-FEB-1999
LOCUS      Human papillomavirus type 9 genomic DNA.
DEFINITION      X74464
ACCESSION      G397068
KEYWORDS      X74464.1      GI:397068
SOURCE      E1 gene; E2 gene; E4 gene; E6 gene; E7 gene; early protein; L1
            gene; L2 gene; late protein.
ORGANISM      Human papillomavirus type 9.
            Human papillomavirus type 9
            Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
            Papillomavirus.
REFERENCE      1 (bases 1 to 7434)
AUTHORS      Dellus,H.
TITLE      Direct Submission
JOURNAL      Submitted (06-AUG-1993) H. Dellus, Deutsches
            Krebsforschungszentrum, Abteiling ATV, Im Neuenheimer Feld 506, W
            6900 Heidelberg, FRG
REFERENCE      2 (bases 1 to 7434)
AUTHORS      Dellus,H. and Hofmann,B.
TITLE      Primer-directed sequencing of human papillomavirus types
JOURNAL      Curr. Top. Microbiol. Immunol. 186, 13-31 (1994)
FEATURES
SOURCE      Location/Qualifiers
            1..7434
            /organism="Human papillomavirus type 9"
            /db_xref="taxon:10621"
            /clone="insert in BamHI site of pBR322"
            200..646
            /gene="E6"
            200..646
            CDS

```

gene  
CDS  
/gene="E6"  
/note="alternative"  
/codon\_start=1  
/product="early protein"  
/protein\_id="CA52482.1"  
/db\_xref="PID:g397069"  
/db\_xref="GI:397069"  
/db\_xref="SWISS-PROT:P36801"  
/translation="MYTLEIOLMDRPRRTKELADLTIVPLIDLLIPCKRCNFIISYF  
ELNFDNRKQDLITTEEDLYVGLCCSCAYASAOLETFHFQFPAVNGKDITVEGTGTAIG  
NICRCRCFELDLVERLATCYKFEQFYKRNMSKGLCRHCGSE"  
221..646  
/gene="E6"  
/codon\_start=1  
/product="early protein"  
/protein\_id="CA52483.1"  
/db\_xref="PID:g397070"  
/db\_xref="GI:397070"  
/db\_xref="SWISS-PROT:P36801"  
/translation="MDRKPRTVELADLTIVPLIDLLIPCKRCNFIISYFELNFDNRKQDLITTEEDLYVGLCCSCAYASAOLETFHFQFPAVNGKDITVEGTGTAIG  
NICRCRCFELDLVERLATCYKFEQFYKRNMSKGLCRHCGSE"  
643..924  
/gene="E7"  
/codon\_start=1  
/product="early protein"  
/protein\_id="CA52484.1"  
/db\_xref="PID:g397071"  
/db\_xref="GI:397071"  
/db\_xref="SWISS-PROT:P36817"  
/translation="WIGKEATPEVLELEOVLQPTADLHCYEELTEPRPEEOCLTP  
YKIVAGCCGARRLVATNLGIRAOELLGLDVLQVCECKRLNHE"  
917..2734  
/gene="E1"  
917..2734  
/gene="E1"  
/codon\_start=1  
/product="early protein"  
/protein\_id="CA52485.1"  
/db\_xref="PID:g397072"  
/db\_xref="GI:397072"  
/db\_xref="SWISS-PROT:Q05111"  
/translation="WSDNRKTKPDLKKECCSNWLSLEACSDSSLDGDELKLPFBGSDS  
DISLIDIDGDVQGNRSRLFCQDESEBOOTOLKRRYISPAVVLQLSQLEISISLS  
POHKRRALFEQDGLSECVNEADLSETQVEVPANPTTAQTKGLGIKDLKLS  
NKAIVLAKKEAFEGVFAELTRYKSNKTCGRCBMVAAYAVNDLLESSKOLLQHC  
AYIMHYHPMCIYLCITCNKGRSREYCRLLSTLQVSEOLISEPRLTSCVAALEP  
IKGNSNPVAVHGAIPENILQTLINHOANATFDSLIMQFIYDHEIYDEATIAQ  
YAKLEITANRAFLQNSQARLVKCAATWRHATRGEMEMSWIIRKLLLVESN  
GOMSVIPIRYIQDINFLEELTFKFAELONKPKONCLFPGPDGSKMTMSLIIVL  
KGKVIYFNGSTFEMLOIADTKLALIDVYHVCWEYIDQYLRNGLDGNYCGLDMKRR  
APCKMRPPLNLSTNIDITKQKRYLHRSKPAFNNKPLDANHKPOEELTDQSWK  
SEFKRLTQQLDSQEDGEDGNSQRTFCIARDFNGPV"  
2676..4061  
/gene="E2"  
2676..4061  
/gene="E2"  
/codon\_start=1  
/product="early protein"  
/protein\_id="CA52486.1"  
/db\_xref="PID:g397073"  
/db\_xref="GI:397073"  
/db\_xref="SWISS-PROT:P36780"  
/translation="METLSARFNALQETIMDIYESGREDDQSOIDHMOQLROBQILH  
YAKRQVRLGQPVPLATSEQAKDAIGVLLQSLQSRAYQGEPTTIAQISLEAY  
RSPPVARKKGPQNIIEVYDDPDNVMSYITWNFIYQVAVDWEKVOGVHDFGAY  
REGVATLYINFDKDAARYGRTGWENVDIVFAVYSSPTGOGGETSKRTLYR  
SSGPTSTPLATVYPTGSGTSRRYORKASSPTRRKROGQEGEGEEETNRYR  
QSRSGAGTNERGGERGRRGSSADSTPTDGRRLARRLLAEKDDPLMLGDNAVSKC  
RRSGGKATRGVSGVPEVGTGVRVAGHHGRIARRLLAEKDDPLMLGDNAVSKC  
YFRRRRKKRGILVYKYSTTWSVGEDSCDRGRAMTLADPTYIHRQOFTIRTKLPTP





RESULT	11			PRI	03-OCT-1995
HUMNEURO/c					
LOCUS	HUMNEUROF_100849	bp	DNA		
DEFINITION	Human oligodendrocyte myelin glycoprotein (OMG) exons 1-2; neurofilament protein 1 (NFL) exons 28-49; ecotropic viral integration site 2B (EV12B) exons 1-2; ecotropic viral integration site 2A (EV12A) exons 1-2; adenylate kinase (AK3) exons 1-2.				
ACCESSION	L05367	L03723			
NID	G189152				
VERSION	L05367.1	G1:189152			
KEYWORDS	adenylate kinase; ecotropic viral integration site 2A; ecotropic viral integration site 2B; neurofilament protein type 1; oligodendrocyte myelin glycoprotein.				
SOURCE	Homo sapiens DNA.				
ORGANISM	Homo sapiens				
	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 100849)				
AUTHORS	Cawthon,R.M., Weiss,R.B., Xu,G., Viskochil,D., Culver,M., Stevens,J., Robertson,M., Dunn,D., Gesteland,R., O'Connell,P. and White,R.				
TITLE	A major segment of the neurofibromatosis type 1 gene, cDNA sequence, genomic structure, and point mutations [published erratum appears in Cell 1990 Aug 10;62(3):following 608]				
JOURNAL	Cell	62	(1),	193-201	(1990)
MEDLINE	90304909				
REFERENCE	2 (bases 1 to 100849)				
AUTHORS	Cawthon,R.M., O'Connell,P., Buchberg,A.M., Viskochil,D., Weiss,R.B., Culver,M., Stevens,J., Jenkins,N.A., Copeland,N.G. and White,R.				
TITLE	Identification and characterization of transcripts from the neurofibromatosis 1 region: the sequence and genomic structure of EV12 and mapping of other transcripts				
JOURNAL	Genomics	7	(4),	555-565	(1990)
MEDLINE	90353953				
REFERENCE	3 (bases 1 to 100849)				
AUTHORS	Cawthon,R.M., Andersen,L.B., Buchberg,A.M., Xu,G.F., O'Connell,P., Viskochil,D., Weiss,R.B., Wallace,M.R., Marchuk,D.A., Culver,M., Stevens,J., Jenkins,N.A., Copeland,N.G., Collins,F.S. and White,R.				
TITLE	cDNA sequence and genomic structure of EV12B, a gene lying within an intron of the neurofibromatosis type 1 gene				
JOURNAL	Genomics	9	(3),	446-460	(1991)
MEDLINE	9126164				
REFERENCE	4 (bases 1 to 100849)				
AUTHORS	Viskochil,D., Cawthon,R., O'Connell,P., Xu,G.F., Stevens,J., Culver,M., Carey,J. and White,R.				
TITLE	The gene encoding the oligodendrocyte-myelin glycoprotein is embedded within the neurofibromatosis type 1 gene .				
JOURNAL	Mol. Cell. Biol.	11	(2),	906-912	(1991)
MEDLINE	9117257				
REFERENCE	5 (bases 1 to 100849)				
AUTHORS	Wallace,M.R., Andersen,L., Letcher,R., Odell,H., Saulino,A.M., Fountain,D., Breiten,A., Nicholson,J., Mitchell,A.,				

addressing Dr. Robert Weiss may be contacted at the following address and telephone number:

	e-mail:
bob@corona.med.utah.edu	post-al: Room
7160, Eccles Institute of Human Genetics, University of Utah, Salt Lake City, Utah 84112	telephone: (801) 585-3436.
Location/Qualifiers	

LOCATIONS	Location/Qualifiers
source	1. .100849

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="17q11.2"
<1. 36919
intron

```

gene	complement(5761. .8342,
exon	/gene="OMG" complement(5761. .7447)

```

/number=2
complement(7448. .8265)

```

exon	complement(8266. .8342) /gene="OVC"
------	--

**repeat\_region**

```

    /rpt_family="GT"
    complément(14658. .16737)
    /gene="EVI2B"
    /note="G00-125-901"
    /citation={3}

```

gene	complement(14658. .25133)
intron	/gene="EVI2B"
	complement(16738. .25076)
	/gene="EVI2B"

repeat\_region

```

exon
"
/rpt_family="GT"
/segment(25077..25133)
/gene="EVI2b"
/note="G00-125-901"
/citation={3}
/number=1
repeat_region
28465..28503

```

exon

```

/gene="EV12A"
/note="G00-125-191"
/citation="[2]"
/number=2
complement(28542..32788)
/gene="EV12A"

```

**Intron**

intron	complement(30122..	.32588
	/gene="EV12A"	
	/note="G00-125-191"	
exon	/number=1	
	complement(32589..	.32788)

repeat\_region  
exon

gene . . .  
36920. .86010  
/number=28  
CDNA (HUMNF1B,M89914); G00-120-231"  
/nuc= 36920 = nt.5032 from full-length neurofibromin

## Inter

/note="G00-120-231"  
/number=28  
38597. .38937  
exon

```

/note="G00-120-231"
/number=29
38938. .41643
Intron

```

```
/note="G00-120-231"
/number=29
41644. .41846
```

```

/note="G00-120-231"
/number=30
41847. .46170

```

```
/note="G00-120-231"  
/number=30  
46171. .46364
```

```

/note="G00-120-231"
/number=31
replace(46216,"c")
utation

```

```

/note="leu to pro change, exon 31; G00-120-231"
/citation=[1]
replace(46260,"t")
utation

```

```

/note="arg to stop change, exon 31; G00-120-231"
/citation=[1]
46365..47772
nttron

```

```

/note="G00-120-231"
/number=31
replace(46600. .46740,"")

```

translocation breakpoint; G00-120-231" /note="CI/122 47773. .47913 kon

```

/note="G00-120-231"
/number=32
47914. .48075

```

```

/note="G00-120-231"
/number=32
replace(48030,"")
itation

```

```
/note="de novo ALU insertion; G00-120-231"  
/citation=[5]  
48076.48355
```

8 TAGTTCGAATTGCTAATCCAGCAATTTGTGGATAGCTGCAAACTGCCATATGTA 60  
||||| | ||||| ||| ||||| | |

ORGANISM	REFERENCE
<i>Homo sapiens</i>	1 (Dases 1 to 297898)
Euthariota; Metazoa; Chordata; Cranialata; Mammalia;	
Eutharia; Primates; Catarrhini; Homnidae; Homo.	
WEISS, R. B., DUNN, D. M., AYOGLU, A., BANKS, L., DISERA, L., DUVAL, B.	

Hamill, C., Holmes, C., Korzenowski, G., Mahmoud, M., Rose, R., Stokes, R., Stump, D., Yu, P., Zhou, L., Giffin, Y., Nelson, J. and

```

JOURNAL      vonNiederhausen,A.
REFERENCE    Unpublished
AUTHORS      2 (bases 1 to 297898)
              Weiss,R.B., Dunn,D.M., Aoyagi,A., Banks,L., Disera,L., Duval,B.,
              Hamil,C., Holmes,C., Kozienowski,G., Mahmoud,M., Rose,R.,
              Stokes,R., Wang,C., Yu,P., Zhou,L., Giltin,Y., Nelson,J. and
              vonNiederhausen,A.
TITLE        Direct Submission
JOURNAL      Submitted (04-APR-1998) Human Genetics, University of Utah, 20 S.
REFERENCE    2030 E., Rm 308, Salt Lake City, Utah 84112, U.S.A.
AUTHORS      3 (bases 1 to 297898)
              Weiss,R.B., Dunn,D.M., Aoyagi,A., Banks,L., Disera,L., Duval,B.,
              Hamil,C., Holmes,C., Mahmoud,M., Rose,R., Stokes,R., Stump,D.,
              Yu,P., Zhou,L., Giltin,Y., Nelson,J. and vonNiederhausen,A.
TITLE        Direct Submission
JOURNAL      Submitted (17-OCT-1998) Human Genetics, University of Utah, 20 S.
REFERENCE    2030 E., Rm 308, Salt Lake City, Utah 84112, U.S.A.
AUTHORS      4 (bases 1 to 297898)
              Weiss,R.B., Dunn,D.M., Aoyagi,A., Banks,L., Disera,L., Duval,B.,
              Hamil,C., Holmes,C., Mahmoud,M., Rose,R., Stokes,R., Stump,D.,
              Yu,P., Zhou,L., Giltin,Y., Nelson,J. and vonNiederhausen,A.
TITLE        Direct Submission
JOURNAL      Submitted (22-OCT-1998) Human Genetics, University of Utah, 20 S.
REFERENCE    2030 E., Rm 308, Salt Lake City, Utah 84112, U.S.A.
AUTHORS      5 (bases 1 to 297898)
              Weiss,R.B., Dunn,D.M., Aoyagi,A., Banks,L., Disera,L., Duval,B.,
              Hamil,C., Holmes,C., Mahmoud,M., Rose,R., Stokes,R., Stump,D.,
              Yu,P., Zhou,L., Giltin,Y., Nelson,J. and vonNiederhausen,A.
TITLE        Direct Submission
JOURNAL      Submitted (25-FEB-1999) Human Genetics, University of Utah, 20 S.
REFERENCE    2030 E., Rm 308, Salt Lake City, Utah 84112, U.S.A.
AUTHORS      On Oct 22, 1998 this sequence version replaced g1:3766112.
COMMENT      Location/Qualifiers
FEATURES      1..297898
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="23013,14502"
              /chromosome="21"
              /clone_id="CITB-978SK-B"
BASE COUNT    82446 a 60770 c 62692 g 91988 t      2 others
ORIGIN
Query Match   38.5%; Score 25.8; DB 42; Length 297898;
Best Local, Similarity 63.9%; Pred. No. 19;
Matches 39; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
1 TTTATCTAGTTCATTCATTCAGCATTTGTGATAGCTGCAACCTGCGATATGTA 60
188378 TCTTATTAGTGTATCTTATCTTATCTGCGAATTAAGGATTAACGACGACTGTCTATTTC 188319
QY      61 A 61
Db 188318 A 188318

```

```

AUTHORS      Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (12-JUN-1998) Genome Sequencing Center, Washington
REFERENCE    University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
AUTHORS      3 (bases 1 to 141313)
              Waterston,R.
TITLE        Direct Submission
JOURNAL      Submitted (24-NOV-1998) Department of Genetics, Washington
REFERENCE    University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
              On Nov 24, 1998 this sequence version replaced g1:3264783.
COMMENT      SUBMITTED BY: WUGSC
              Genome Sequencing Center
              Department of Genetics
              Washington University
              St. Louis MO 63108, USA
              http://genome.wustl.edu/gsc
              mailto:saplens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-5, prepared by
Pleier de Jong and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu) using the method described by
Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from
one male donor.
The clone may be obtained either from Genome Systems, Inc.
(http://www.genomesystems.com) or Research Genetics, Inc.
(http://www.resgen.com); or from Pleier de Jong.
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:
Actual start of this clone is at base position 1 of DJ0991G20;
actual end is at 141313 of DJ0991G20

The location of this clone is unknown.
FEATURES      Location/Qualifiers
              1..141313
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="unknown"
              /clone_id="RPCI-5"
              /clone="DJ0991G20"
              30..152
              /rpt_family="Retroviral"
              405..634
              /rpt_family="MIR"
              686..865
              /rpt_family="MER2_type"
              866..1014
              /rpt_family="MIR"
              1047..1185
              /rpt_family="MIR"
              1347..1969
              /rpt_family="AT-rich"
              3058..3184
              /rpt_family="Alu"
              3447..4731
              /note="cytochrome pseudogene"
              4869..5017
              /rpt_family="L2"
              5039..5247
              /rpt_family="MIR"

```

```

repeat_region      5283. .5467      /rpt_family="L2"
repeat_region      5512. .11701    /rpt_family="L1"
repeat_region      11826. .12139    /rpt_family="Alu"
repeat_region      12557. .12585    /rpt_family="Alu"
repeat_region      12586. .12704    /rpt_family="BC200"
repeat_region      13614. .13642    /rpt_family="Alu"
repeat_region      13655. .13952    /rpt_family="AT_rich"
repeat_region      14871. .14956    /rpt_family="Alu"
repeat_region      14973. .15007    /rpt_family="L2"
repeat_region      15010. .15149    /rpt_family="AT_rich"
repeat_region      15159. .15429    /rpt_family="Alu"
repeat_region      15850. .15984    /rpt_family="L2"
repeat_region      16009. .16299    /rpt_family="MIR"
repeat_region      16840. .16947    /rpt_family="Alu"
repeat_region      17292. .17343    /rpt_family="purine-rich"
repeat_region      17684. .18442    /rpt_family="L2"
repeat_region      18523. .18544    /rpt_family="L1"
repeat_region      19220. .19245    /rpt_family="AT_rich"
repeat_region      20355. .20789    /rpt_family="AT_rich"
repeat_region      20862. .20991    /rpt_family="L1"
repeat_region      21053. .21172    /rpt_family="MER94"
repeat_region      21409. .21712    /rpt_family="MER1_type"
repeat_region      22318. .22613    /rpt_family="Alu"
repeat_region      23077. .23206    /rpt_family="Alu"
repeat_region      24150. .24824    /rpt_family="Alu"
repeat_region      25528. .25966    /rpt_family="L1"
repeat_region      26005. .26113    /rpt_family="Retroviral"
repeat_region      26125. .26800    /rpt_family="Alu"
repeat_region      26801. .27101    /rpt_family="Retroviral"
repeat_region      27102. .27431    /rpt_family="Alu"
repeat_region      27433. .27560    /rpt_family="Retroviral"
repeat_region      27643. .27725    /rpt_family="L1"
repeat_region      27726. .28016    /rpt_family="Alu"
repeat_region      28017. .28510    /rpt_family="Alu"
repeat_region      28542. .28617    /rpt_family="L1"
repeat_region      28676. .29224    /rpt_family="AT_rich"

```

```

repeat_region      29558. .29851    /rpt_family="MALR"
repeat_region      30066. .30242    /rpt_family="Alu"
repeat_region      30469. .30764    /rpt_family="MALR"
repeat_region      30856. .31238    /rpt_family="Alu"
repeat_region      31239. .31306    /rpt_family="L1"
repeat_region      31308. .31970    /rpt_family="(CAT)n"
repeat_region      31971. .32271    /rpt_family="L1"
repeat_region      32272. .32324    /rpt_family="Alu"
repeat_region      32344. .32683    /rpt_family="L1"
repeat_region      32659. .32721    /rpt_family="L1"
repeat_region      32741. .32865    /rpt_family="L1"
repeat_region      32876. .34100    /rpt_family="Alu"
repeat_region      34101. .34374    /rpt_family="L1"
repeat_region      34375. .35747    /rpt_family="Alu"
repeat_region      35735. .36153    /rpt_family="L1"
repeat_region      36154. .36701    /rpt_family="MALR"
repeat_region      36852. .37128    /rpt_family="MALR"
repeat_region      37644. .38038    /rpt_family="L1"
repeat_region      40126. .40182    /rpt_family="MALR"
repeat_region      /rpt_family="Alu"

```

```

Query Match      38.2%; Score 25.6; DB 11; Length 141313;
Best Local Similarity 66.1%; Pred. No. 24;
Matches 37; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY      1 TTATTCCTAGTCCATTCCTAATCCAGCATTTGTGATAGCTGCATACTGGATA 56
Db 90935 TTTCCTCTAGTCTTTTCTTCTAATACCTACTTTTACATAGCTGTACATGGATA 90880

```

Search completed: September 17, 1999, 22:04:19  
 Job time: 16442 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:20 ; Search time 2825.05 Seconds

(without alignments)  
46.781 Million cell updates/sec

Title: US-08-956-518a-98

Sequence: 1 TTTTCTAGTTCAGTTC.....ACTGCGATGCTAGTACCA 67

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: em\_est20: \*  
21: em\_est21: \*  
22: em\_est22: \*  
23: em\_est23: \*  
24: em\_est24: \*  
25: em\_est25: \*  
26: em\_est26: \*  
27: em\_est27: \*  
28: em\_est28: \*  
29: em\_est29: \*  
30: em\_est30: \*  
31: em\_est31: \*  
32: em\_est32: \*  
33: em\_est33: \*  
34: em\_est34: \*  
35: em\_est35: \*  
36: em\_est36: \*  
37: em\_est37: \*  
38: em\_est38: \*  
39: em\_est39: \*  
40: em\_est40: \*  
41: em\_est41: \*  
42: em\_est42: \*  
43: em\_est43: \*  
44: em\_est44: \*  
45: em\_est45: \*  
46: em\_est46: \*  
47: em\_est47: \*  
48: em\_est48: \*  
49: em\_est49: \*  
50: em\_est50: \*  
51: em\_est51: \*  
52: em\_est52: \*  
53: em\_est53: \*

54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	48.6	72.5	487	39	AA861176	AA861176 AK36409.s
2	48.2	71.9	560	26	W52861	W52861 zc03e02.r1
3	44.4	65.7	409	25	W03952	W03952 za62c08.r1
4	43.2	64.5	274	34	AA460018	AA460018 z49h09.s
5	43.2	64.5	436	39	AA844642	AA844642 a357a12.s
6	43.2	64.5	396	45	AI367541	AI367541 qy93c12.x
7	43.2	64.5	501	49	AI655990	AI655990 tt42c03.x
8	42.7	40.3	354	26	W85937	W85937 zh60e03.r1
9	26.6	39.7	560	26	W76900	W76900 me58b02.r1
10	26.6	39.7	416	26	W85213	W85213 m51b09.r1
11	26.6	39.7	517	27	W89656	W89656 m81f04.r1
12	26.6	39.7	590	30	AA222672	AA222672 mv99e05.r
13	26.6	39.7	530	33	AA388232	AA388232 vb59c01.r
14	26.6	39.7	545	35	AA590415	AA590415 vml6d07.r
15	26.6	39.7	479	37	AA719305	AA719305 zb35c02.s
16	26.6	39.7	499	38	AA771094	AA771094 vt16e07.r
17	26.6	39.7	542	38	AA792253	AA792253 vn94e09.r
18	26.6	39.7	618	40	C89142	C89142 C89142 Mous
19	26.6	39.7	1048	44	AI255545	AI255545 u155h01.y
20	25.6	38.2	291	20	T05021	T05021 EST02909.Fe
21	25.2	37.6	445	47	AI488647	AI488647 EST246986
22	25.2	37.6	451	47	AI488637	AI488637 EST247176
23	24.6	36.7	467	42	AI099847	AI099847 34000.Lam
24	24.6	36.7	347	43	C99808	C99808 C99808 YAC
25	24.6	36.7	297	48	AI555647	AI555647 UI-R-C2P-
26	24.4	36.4	633	25	W10059	W10059 me67d04.r1
27	24.4	36.4	610	27	AA027731	AA027731 m14e01.r
28	24.4	36.4	669	27	AA027739	AA027739 m14g01.r
29	24.4	36.4	611	27	AA033365	AA033365 m143d06.r
30	24.4	36.4	582	27	AA048492	AA048492 m30c08.r
31	24.4	36.4	627	28	AA061156	AA061156 m31c08.r
32	24.4	36.4	466	28	AA103275	AA103275 mo25h04.r
33	24.4	36.4	420	28	AA107857	AA107857 mo49h08.r
34	24.4	36.4	322	28	D76888	D76888 MDS91F01.mo
35	24.4	36.4	654	29	AA170084	AA170084 ms32g06.r
36	24.4	36.4	593	29	AA183839	AA183839 mo95a04.r
37	24.4	36.4	576	29	AA184588	AA184588 mt51c02.r
38	24.4	36.4	498	30	AA197755	AA197755 mv03a06.r
39	24.4	36.4	490	34	AA517605	AA517605 v933d04.r
40	24.4	36.4	616	35	AA544863	AA544863 vk38b07.r
41	24.4	36.4	500	35	AA561216	AA561216 v134e02.r
42	24.4	36.4	554	37	AA711653	AA711653 vu27b09.r
43	24.4	36.4	828	38	AA755792	AA755792 vv35d03.r
44	24.4	36.4	555	38	AA794314	AA794314 vu77e08.r
45	24.4	36.4	880	50	AU066965	AU066965 AU066965

## ALIGNMENTS

RESULT 1  
AA861176/c  
LOCUS  
DEFINITION AK36409.s1 Soares\_testis\_NHT Homo sapiens cDNA clone IMAGE:1408049  
ACCESSION AA861176  
NID 92953316  
VERSION AA861176.1 GI:2953316







Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 415.  
Location/Qualifiers  
1. 436

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="1; 1"  
/clone="IMAGE:1394398"  
/clone.lib="Soares\_testis\_NH"  
/sex="male"  
/lab\_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from Clontech  
Laboratories, Inc., and primed with a Not I - oligo(dT)  
primer [5']  
TGTTCACATTCGATCAGACGAGCGGCCGCCCAATTTTCTTTTCTTTT 3']  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization to Cots5, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 115 a 96 c 103 g 122 t

ORIGIN

Query Match 64.5%; Score 43.2; DB 39; Length 436;  
Best Local Similarity 93.8%; Pred. No. 2.2e-05;  
Matches 45; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 10 GTTCCATTCGATCAGACGATTTGTGATAGCTGCAACTGGCATAT 57  
|||||  
Db 267 GTTCCATTCGATCAGACGATTTGTGATAGCTGCAACTGGCATAT 220

RESULT 6  
A1367541/c 396 bp mRNA EST 15-FEB-1999  
LOCUS qv93c12.x1 NCI\_CGAP\_Ut2 Homo sapiens cDNA clone IMAGE:1989142 3',  
DEFINITION mRNA sequence.  
ACCESSION A1367541  
NID 94137286  
VERSION A1367541.1 GI:4137286  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 396)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)

JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www.bio.lnl.gov/bdrip/image/image.html

FEATURES  
source  
Insert Length: 1828 Std Error: 0.00  
Seq primer: -40UP from GIBCO  
High quality sequence stop: 370.  
Location/Qualifiers  
1. 396  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

/clone="IMAGE:1989142"  
/clone.lib="NCI\_CGAP\_Ut2"  
/tissue\_type="moderately-differentiated endometrial  
adenocarcinoma, 3 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Clone: unidirectionally. Primer: Oligo dT.  
Average insert size 1.85 kb. Life Technologies catalog #:  
11539-012"

BASE COUNT 98 a 86 c 97 g 115 t

ORIGIN

Query Match 64.5%; Score 43.2; DB 45; Length 396;  
Best Local Similarity 93.8%; Pred. No. 2.2e-05;  
Matches 45; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 10 GTTCCATTCGATCAGACGATTTGTGATAGCTGCAACTGGCATAT 57  
|||||  
Db 278 GTTCCATTCGATCAGACGATTTGTGATAGCTGCAACTGGCATAT 231

RESULT 7  
A1655990/c 501 bp mRNA EST 04-MAY-1999  
LOCUS t142c03.x1 NCI\_CGAP\_GC6 Homo sapiens cDNA clone IMAGE:2243428 3',  
DEFINITION mRNA sequence.  
ACCESSION A1655990  
NID 94739969  
VERSION A1655990.1 GI:4739969  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 501)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Jun 5, 1998 this sequence version replaced gi:3188057.

JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www.bio.lnl.gov/bdrip/image/image.html

FEATURES  
source  
Seq primer: -40UP from GIBCO  
High quality sequence stop: 458.  
Location/Qualifiers  
1. 501  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2243428"  
/clone.lib="NCI\_CGAP\_GC6"  
/tissue\_type="pooled germ cell tumors"  
/lab\_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Plasmid DNA from the normalized library  
NCI\_CGAP\_GC4 was prepared, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from a pool of 5,000 clones made  
from the same library (clones IDs 1257096-1258631,  
1469064-1470983, and 1475592-1476743). Subtraction by



Query Match	39.7%	Score 26.6;	DB 26;	Length 416;
Best Local Similarity	71.4%;	Pred. No. 10;		
Matches 35;	Conservative	0;	Mismatches 14;	Indels 0;
				Gaps 0;

QY 18 TCGTAAATCCAGCAATTTGTGATAGCTGCAAACTGGATATATCAATAAC 66  
||||| ||| ||||| ||| | ||| ||||| ||  
Db 82 TCGTAAATCCATACATTTTGTGATAGCAGCCAAAGAACTATGTGGAAATGAC 130

RESULT	11
W89656	
LOCUS	
DEFINITION	m18i104.r1 Soares mouse embryo NME1 c 14 c wnt
	W89656 517 bp mRNA
	EST 12-SEP-1996

clone IMAGE:420703 5' similar to gb:U36962 Mus musculus cDNA  
mRNA, complete cds (MOUSE); mRNA sequence.  
W89656

```

NID          91542182
VERSION      W89656.1  GI:1542182
KEYWORDS
SOURCE       EST.
             house mouse.

```

ORGANISM	Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	1 (bases 1 to 517)
ALTIMORE	

**MEMBERS**  
Marra, M., Kuller, L., Allen, M., Bowles, N., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, R., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterson, R.

**TITLE**  
The WashU-HMI Mouse EST Project

**JOURNAL**  
Unpublished (1996)

On Sep 17, 1996 this sequence version replaced gl:1404980.

CONTACT:

Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu

This clone is available royalty-free through LNC; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:255235

Seq primer: -28M13 rev2 from Amer sham  
High quality sequence stop: 492.  
Location/Qualifiers

FEATURES

```

1. 517
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="3 p24.2-p21.1"
/clone_image="420703"
/clone_id="Soares mouse embryo NDMEl3.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10b"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTACCACTCATGAGTCGGAGCGGCGGGAATTTTTTTTTTTTTTTTTTTT
T 3']; on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Mhoru Ro, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and
M.Fatima Bonaldo. "

```

DB	LOCUS	DEFINITION	ACCESSION	NID	VERSION	KEYWORDS	TRACE	ORGANISM	TITLE	JOURNAL	COMMENT
0Y	18	TGCTAATCCAGCATTTGGATAGCTGCAACATCGCATATGTAAGTAC	66								
DB	318	TGCTAATCCAGCATTTGGATAGCTGCAACATCGCATATGTAAGTAC	366								
Query Match	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB 30;	Length 590;								
Best Local	Similarity	71.4%	Pred. No. 10;	Matches	35;	Conservative	0;	Mismatches	14;	Indels	0;
Score	26.6;	DB									

```

RESULT 13
AA388232
LOCUS
DEFINITION
AA388232
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AA388232 530 bp mRNA EST 23-APR-1997
V596C01.1 Ko mouse embryo 11 5dpc Mus musculus cDNA clone
IMAGE:761280 5' similar to gb:U36962 Mus musculus metaxin mRNA,
complete cds (MOUSE);, mRNA sequence.

AA388232
g2041187
AA388232.1 GI:2041187
EST.
house mouse.
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murineae; Mus.

1 (pages 1 to 530)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Therrien, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1407058.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:462200
High quality sequence stop: 445.
Location/Qualifiers
1..530
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:761280"
/clone_11b="Ko mouse embryo 11 5dpc"
/sex="pooled"
/tissue="embryo"
/dev_stage="11.5dpc"
/lab_host="DH10B"
/Note="Organ: embryo; Vector: pSPORT1; Site_1: SalI;
Site_2: NotI; Total RNAs were extracted from 11.5 dpc
embryos (excluding placenta and yolk sac). The
double-stranded cDNA was synthesized with an oligo (dT)-1
primer GAGAGACACTGATCTGATCGGACGCGCGCTTTTTTTTTTTTTT
3'. The cDNAs were ligated to Lt-Sal3a: 5'
GGTATGACGTCGACATAC 3' and Lt-Sal3b: 5'
GGATGTCGACGTCAT 3'. The cDNAs were size-selected and
amplified by long-range PCR using Ex taq polymerase for 10
cycles. The PCR-amplifiable cDNA mixture went through
one round of equalization and was digested with SalI/NotI
and cloned into the SalI/NotI sites of the pSPORT1
plasmid vector (Life Technologies). The library was
constructed by Dr. Minoru S. H. Ko and Dr. Xiaohong
Wang."
BASE COUNT 134 a 140 c 143 g 113 t
ORIGIN
Query Match 39.7%; Score 26.6; DB 33; Length 530;
Best Local Similarity 71.4%; Pred. No. 10;
Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
18 TGGTAAATCCAGCATTTGGATAGCTGCAAACTCGATATGTAGTAATAC 66

```



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:04:19 ; Search time 1962.17 Seconds  
(without alignments)  
162,081 Million cell updates/sec

Title: US-08-956-518a-99  
Perfect score: 100  
Sequence: 1 CTGCTTCTAGTGTGCTGATGAG.....CTGCCTCCAGGTAGCTGCA 100

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

ibase :

GenEmbl: \*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_st:\*  
14: gb\_sy:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_vl:\*  
18: em\_fun:\*  
19: em\_hlg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_hn:\*  
23: em\_om:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_ro:\*  
30: em\_st:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_vl:\*  
34: gb\_hlg1:\*  
35: gb\_hlg2:\*  
36: gb\_hn1:\*  
37: gb\_hn2:\*  
38: em\_ba1:\*  
39: em\_ba2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pr4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result Queue Match Length DB ID Description  
No. Score Query  
1 83.4 83.4 1559 9 HUMA7NAR L25827 Human a7 nt

2	83.4	83.4	2087	10	HSARA7A	X70297 H.sapiens m
3	83.4	83.4	1555	10	HSCHRNA7A	Z23141 H.sapiens C
4	83.4	83.4	1509	10	HSNACHRA7	Y08420 H.sapiens m
5	83.4	83.4	1976	10	HSU62436	U62436 Human nicot
6	82.8	82.8	641	11	AF029838	AF029838 Homo sapi
7	82.8	82.8	689	11	AF029839	AF029839 Homo sapi
8	82.8	82.8	1712	11	AF036903	AF036903 Homo sapi
9	82.8	82.8	1836	11	AF037646	AF037646 Homo sapi
10	81.8	81.8	1977	11	HSU40583	U40583 Human alpha
11	69.0	69.0	1848	12	M05NARS	L37663 Mus musculus
12	69.0	69.0	2106	12	RATNARAD	L31619 Rattus ratt
13	69.0	69.0	3030	12	S53987	S53987 nicotinic r
14	67.4	67.4	1551	3	BTA7	X93604 B.taurus nr
15	57.8	57.8	2696	4	GGATNAREC	X52295 Chicken alp
16	57.8	57.8	2037	4	GGNARA7A	X68586 G.gallus nr
17	48.8	48.8	2030	4	GGANAREC	X52286 Chicken alp
18	48.2	48.2	336	14	G47584	G47584 Z25357.1 Ze
19	42.8	42.8	3029	37	AF143847	AF143847 Heliothis
20	40.0	40.0	89871	37	AC005890	AC005890 Drosophila
21	35.4	35.4	76947	36	AC004326	AC004326 Drosophila
22	33.6	33.6	143664	35	AC007291	AC007291 Drosophila
23	32.6	32.6	1008	12	RATNACHRS	M33952 Rat neurona
24	31.0	31.0	41370	37	CEL30612	U21319 Caenorhabdi
25	30.6	30.6	2493	12	RNNIACETE	X15834 R.novegicu
26	30.6	30.6	2461	12	RNT42976	U42976 Rattus norv
27	30.4	30.4	108487	10	HS738915	AL035252 Human DNA
28	29.6	29.6	2682	3	GOIWTGRC	M55541 Capra hircu
29	29.0	29.0	3629	37	AF143846	AF143846 Heliothis
30	28.8	28.8	1609	11	AF045765	AF045765 Homo sapi
31	28.8	28.8	60896	35	AC007582	AC007582 Drosophila
32	28.2	28.2	3841	4	GGU62904	U62904 Gallus gall
33	28.0	28.0	42782	1	BS43KEDNA	AJ223978 Bacillus
34	28.0	28.0	217420	1	BSUB0017	Z89120 Bacillus su
35	28.0	28.0	209510	1	BSUB0018	Z89121 Bacillus su
36	28.0	28.0	4085	4	CHKACHRO2	K02904 Chicken nic
37	28.0	28.0	415	10	PTU11795	U11795 Pan troglod
38	27.8	27.8	3315	12	MMDMRK	Z21504 M.musculus
39	27.8	27.8	12371	12	AC004728	M6494 Boselaphus
40	27.6	27.6	85439	37	AC004728	AC004728 Drosophila
41	27.4	27.4	2685	3	BOSMTGRC	AC006447 Mus muscu
42	27.4	27.4	140186	35	AC006447	AC006447 Mus muscu
43	27.4	27.4	151528	35	AC005557	AC005557 Drosophila
44	27.4	27.4	257867	37	AC005557	AC005557 Drosophila
45	27.0	27.0	339	9	HOMTCRAVE	L06884 Homo sapien

#### ALIGNMENTS

RESULT 1  
LOCUS HUMA7NAR 1559 bp mRNA PRI 18-AUG-1994  
DEFINITION Human a7 nicotinic acetylcholine receptor mRNA.  
ACCESSION L25827  
KEYWORDS G438616  
VERSION L25827.1 GI:438616  
KEYWORDS nicotinic acetylcholine receptor alpha 7 subunit; transmembrane protein.  
SOURCE Homo sapiens (library: Clontech HL1065b; ATCC 37433) fetus brain CDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Dovelette-Stamm, L., Monteggia, L.M., Donnelly-Roberts, D., Wang, M.T., Lee, J., Tian, J. and Giordano, T.  
TITLE Cloning and sequence of the human a7 nicotinic acetylcholine receptor  
JOURNAL Drug Dev. Res. (1993) In press  
FEATURES  
LOCATION/Qualifiers  
1..1559  
/Organism="Homo sapiens"  
/db\_xref="taxon:9606"

/dev\_stage="fetus"  
/tissue\_type="brain"  
/tissue\_lib="Clontech HL1065D; ATCC 37433"  
BASE COUNT 314 a 463 c 440 g 342 t  
ORIGIN

Query Match 83.4%; Score 83.4; DB 9; Length 1559;  
Best Local Similarity 98.8%; Pred. No. 1.6e-20;  
Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 AGTGTGATGAGCGGCTTTCAGCCACATTCACACTACGTTGTGTGTAATCTCTGGG 68  
DB 326 AGTGTGATGAGCGGCTTTCAGCCACATTCACACTACGTTGTGTGTAATCTCTGGG 68  
OY 69 CATTGCCAGTACCTGCTCCAGGTA 93  
DB 386 CATTGCCAGTACCTGCTCCAGGCA 410

..SUIT 2

HSARA7A 2087 bp mRNA PRI 01-JUN-1994  
LOCUS H.sapiens mRNA for neuronal nicotinic acetylcholine receptor  
DEFINITION alpha-7 subunit.  
X70297  
ACCESSION 9496606  
NID X70297.1 GI:496606  
VERSION neuronal nicotinic acetylcholine receptor alpha-7 subunit.  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2087)  
AUTHORS Peng, X., Katz, M., Gerzanich, V., Anand, R. and Lindstrom, J.  
TITLE Human alpha 7 acetylcholine receptor: cloning of the alpha 7  
subunit from the SH-SY5Y cell line and determination of  
pharmacological properties of native receptors and functional alpha  
7 homomers expressed in Xenopus oocytes  
Mol. Pharmacol. 45 (3), 546-554 (1994)  
94195283

JOURNAL MEDLINE  
REFERENCE 2 (bases 1 to 2087)  
AUTHORS Katz, M.  
TITLE Direct Submission  
JOURNAL Submitted (04-FEB-1993) M. Katz, Univ of Pennsylvania School of  
Medicine, Dept of Neuroscience, 38th & Hamilton Walk, Rm 235  
Stemmler Hall, Philadelphia, PA 19104, USA  
LOCATION/Qualifiers

URES  
source  
1. .2087  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="brain"  
/cell\_type="neuroblastoma"  
/cell\_line="SH-SY-5Y"  
/clone\_id="lambda zap II"  
/clone="SHSY3"  
104. .175  
/product="neuronal nicotinic acetylcholine receptor  
alpha-7 subunit"  
104. .1612  
/codon\_start=1  
/product="neuronal nicotinic acetylcholine receptor  
alpha-7 subunit"  
/protein\_id="CAA49778.1"  
/db\_xref="PID:9496607"  
/db\_xref="GI:496607"  
/db\_xref="SWISS-PROT:P36544"  
/translation="MRCSPGVWGLAASLLHVSLOGEPRKIKYELVKNYNLEPRV  
ANDSOPLVFSLQIMDVDERKNVLTNTNHLQMSMDHYLQWVSEPGVKTVP  
PDCQIMKPDILYNSADPERDQFHTNVLVNSGHCQYLPGLFKSSCYIDVWPF  
VOCKIKFGSWSGMSGLDQMGKADISGTPNGEDLVGIPGKRSERFECCKEPY  
DVFTYTMKRRTIYIGLNLIPCVLLSALALVELLPADSGEKISIGITVLLSLTETM  
LVAELMPATSDSVPLIAQVFASTMIYGVVVYTVIYQYHHHDPDGGKMKRWRTVY

CDS

sig\_peptide

LINMCAMFLMKRPGEDKVRPAQCHKORCLASVENSAGVPPASNGNLITIGRGL  
DGVHVPDPSGVYICGRACCPPTDHEHLHGSGPEPDPDAKILEEYRIANRPQC  
DESAVSEKMFACVDRCLTMAFSVFTIICITIGILMSAPNEVAEVSADKPA"  
BASE COUNT 441 a 601 c 573 g 472 t  
ORIGIN  
mat\_peptide 176. .1609

Query Match 83.4%; Score 83.4; DB 10; Length 2087;  
Best Local Similarity 98.8%; Pred. No. 1.6e-20;  
Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 AGTGTGATGAGCGGCTTTCAGCCACATTCACACTACGTTGTGTGTAATCTCTGGG 68  
DB 452 AGTGTGATGAGCGGCTTTCAGCCACATTCACACTACGTTGTGTGTAATCTCTGGG 68  
OY 69 CATTGCCAGTACCTGCTCCAGGTA 93  
DB 512 CATTGCCAGTACCTGCTCCAGGCA 536

RESULT 3

HSCHRNA7A 1555 bp mRNA PRI 03-MAR-1994  
LOCUS H.sapiens CHRNA7 mRNA, 3' end.  
DEFINITION alpha7 subunit.  
X23141  
ACCESSION 9457736  
NID X23141.1 GI:457736  
VERSION alpha7 nicotinic receptor subunit; cholinergic receptor; CHRNA7  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1555)  
AUTHORS Chini, B., Raimond, E., Elgoyhen, A.B., Morali, D., Balzarotti, M. and  
Hellemann, S.  
TITLE Molecular cloning and chromosomal localization of the human alpha  
7 nicotinic receptor subunit gene (CHRNA7)  
Genomics 19 (2), 379-381 (1994)  
94245214  
REFERENCE 2 (bases 1 to 1555)  
AUTHORS Chini, B.  
TITLE Direct Submission  
JOURNAL Submitted (28-JUN-1993) BICE CHINI, CCIPE, Rue de la Cardonille,  
Montpellier, 34094 Cedex, 5, France  
LOCATION/Qualifiers

URES  
source  
1. .1555  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="retina"  
/clone\_id="lambda gt11"  
/clone="CHRNA7"  
1. .1437  
/product="cholinergic receptor, nicotinic, alpha  
polypeptide 7"  
1. .1440  
/gene="CHRNA7"  
/gene="CHRNA7"  
/codon\_start=1  
/product="cholinergic receptor, nicotinic, alpha  
polypeptide 7"  
/protein\_id="CAA80672.1"  
/db\_xref="PID:9457737"  
/db\_xref="GI:457737"  
/db\_xref="SWISS-PROT:P36544"  
/translation="EQRRLYELVKNYNLEPRVANDSOPLVFSLQIMDVDE  
KNQVLTNTNHLQMSMDHYLQWVSEPGVKTVPDQIMKPDILYNSADPERDQ  
FHTNVLVNSGHCQYLPGLFKSSCYIDVWPFVOCCKIKFGSWSGMSGLDQMG  
KADISGTPNGEDLVGIPGKRSERFECCKEPYDVFTYTMKRRTIYIGLNLIP  
CVLLSALALVELLPADSGEKISIGITVLLSLTETM LVAELMPATSDSVPLIAQ  
VFASTMIYGVVVYTVIYQYHHHDPDGGKMKRWRTVY LINMCAMFLMKRPGEDKVRPA

CDS

mat\_peptide

gene



OHKORSSLASVENSANVAPPASNGNLLITGFRGLDGVHCYTPDPSGVGCAMPAPPT  
HDEHLHGOPEDPDPLAKILEEVRITANRCDSEAVCSMKRPACVVDCLM  
AFSVFTICTGILMSAPNVEAVSKDFA"

BASE COUNT 320 a 454 c 434 g 347 t

ORIGIN

Query Match 83.4%; Score 83.4; DB 10; Length 1555;  
Best Local Similarity 98.8%; Pred. No. 1.6e-20;  
Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 AGTGTGATGAGCGCTTTGAGCCGACATTCACACTAAGCTGTGTGATCTCTGGG 68  
DB 280 AGTGTGATGAGCGCTTTGAGCCGACATTCACACTAAGCTGTGTGATCTCTGGG 339  
OY 69 CATTGCCAGTACTGCTCCAGGTA 93  
DB 340 CATTGCCAGTACTGCTCCAGGTA 364

LOCUS 4  
ACCHRA7 1509 bp mRNA PRI 22-JAN-1998  
DEFINITION H.sapiens mRNA for nicotinic acetylcholine receptor alpha7 subunit  
precursor.  
ACCESSION Y08420  
NID Y08420.1 GI:2808623  
KEYWORDS nAChR gene; nicotinic acetylcholine receptor alpha 7 subunit.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 1509)  
AUTHORS Groot Kormelink, P.J. and Luyten, W.H.  
TITLE Cloning and sequence of full-length cDNAs encoding the human  
neuronal nicotinic acetylcholine receptor (nAChR) subunits beta3  
and beta4 and expression of seven nAChR subunits in the human  
neuroblastoma cell line SH-SY5Y and/or IMR-32  
JOURNAL FEBS Lett. 400 (3), 309-314 (1997)  
REFERENCE 2 (bases 1 to 1509)  
AUTHORS Groot Kormelink, P.J.  
TITLE Direct Submision  
JOURNAL Submitted (27-SEP-1996) P.J. Groot Kormelink, Janssen Research  
Foundation, Exp. Mol. Biol. Dept., Turnhoutseweg 30, B-2340 Beerse,  
Belgium

REMARK Revisd by author 22-JAN-1998  
REMARK On Jan 25, 1998 this sequence version replaced gi:1702915.

FEATURES  
SOURCE

1.1509  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="neuroblastoma"  
/cell\_line="IMR-32"  
/cell\_line="SH-SY5Y"  
1.1509  
/gene="nAChRA7"  
1.1509  
/gene="nAChRA7"  
/codon\_start=1  
/product="nicotinic acetylcholine receptor alpha7 subunit  
precursor"  
/protein\_id="CAA69697.1"  
/db\_xref="PID:e1246208"  
/db\_xref="PID:g2808624"  
/db\_xref="GI:2808624"  
/translation="MRCSPGVGLALASLHVSLOGEFORFKLYKELVKNYNPLERPV  
ANDSQPIVYFSLIQLIMDYDEKNOVLTNTNWLQMSWTDHYLQNVSEYPGVKTVEF  
PDGQIKPDLILYNASDERFATFTNVLNNSGHCQLPGLFKSSCTIDVWPFED  
VCHCKLRGNSYSGWSLDLQMDADISGIPNGEWDLVGIPKRSERSEYDECKEPEP  
DVFTVYMRRLTYGILNLLIPCVLISALALVFLPADSGEKISLGITVLISLVFM  
LVAELMPATSDSVPLIAGFASTMTIVGLSVVTVTVLVQYHHHDDPGKMPKMTAVI

CDS

sig\_peptide

gene  
mat\_peptide

BASE COUNT 297 a 451 c 429 g 332 t

ORIGIN

Query Match 83.4%; Score 83.4; DB 10; Length 1509;  
Best Local Similarity 98.8%; Pred. No. 1.6e-20;  
Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 AGTGTGATGAGCGCTTTGAGCCGACATTCACACTAAGCTGTGTGATCTCTGGG 68  
DB 349 AGTGTGATGAGCGCTTTGAGCCGACATTCACACTAAGCTGTGTGATCTCTGGG 408  
OY 69 CATTGCCAGTACTGCTCCAGGTA 93  
DB 409 CATTGCCAGTACTGCTCCAGGTA 433

RESULT 5  
HS062436 1876 bp mRNA PRI 11-JAN-1997  
LOCUS HS062436  
DEFINITION Human nicotinic acetylcholine receptor alpha7 subunit precursor,  
mRNA, complete cds.  
ACCESSION U62436  
NID U62436.1 GI:1458119  
KEYWORDS nAChR gene; nicotinic acetylcholine receptor alpha 7 subunit.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 1876)  
AUTHORS Elliott, K.J., Ellis, S.B., Berckhan, K.J., Urrutia, A.,  
Chavez-Noriega, L.E., Johnson, E.C., Velicic, G., and Harpold, M.M.  
TITLE Comparative structure of human neuronal alpha 2-alpha 7 and beta  
2-beta 4 nicotinic acetylcholine receptor subunits and functional  
expression of the alpha 2, alpha 3, alpha 4, alpha 7, beta 2, and  
beta 4 subunits  
JOURNAL J. Mol. Neurosci. 7 (3), 217-228 (1996)  
REFERENCE 2 (bases 1 to 1876)  
AUTHORS Elliott, K.J.  
TITLE Direct Submision  
JOURNAL Submitted (28-JUN-1996) Kathryn J. Elliott, SIBIA Neurosciences,  
Inc., 505 Coast Blvd. So., La Jolla, CA 92037, USA

FEATURES  
SOURCE

1.1876  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="KBA1pha7.3"  
/clone\_1ib="SIBIA lambda gt11 library #3 (M. Williams)"  
/cell\_line="IMR32"  
/cell\_type="neuroblastoma"  
1.172  
1.138  
73.1581  
/codon\_start=1  
/product="nicotinic acetylcholine receptor alpha7 subunit  
precursor"  
/protein\_id="AAB40114.1"  
/db\_xref="PID:g1458120"  
/db\_xref="GI:1458120"  
/translation="MRCSPGVGLALASLHVSLOGEFORFKLYKELVKNYNPLERPV  
ANDSQPIVYFSLIQLIMDYDEKNOVLTNTNWLQMSWTDHYLQNVSEYPGVKTVEF  
PDGQIKPDLILYNASDERFATFTNVLNNSGHCQLPGLFKSSCTIDVWPFED  
VCHCKLRGNSYSGWSLDLQMDADISGIPNGEWDLVGIPKRSERSEYDECKEPEP  
DVFTVYMRRLTYGILNLLIPCVLISALALVFLPADSGEKISLGITVLISLVFM

CDS

sig\_peptide

5'UTR

## mat\_peptide

LVLAETPATSDEVPILAOYEASTMTITVGLSVVTVIVLYQHHDGDKMKPKMTNVI  
 LLMWCAWFLMKRPGEDKVRAPACOHKORCSLASVEMSAVAPPAENGMLTIGFGL  
 DGVHCVPTDPSGVGCMACSPHDEHLHGGPPGDDPLAKILEVXRINRFQCO  
 DESEAVCSEKRFACVYDRCLMAFVFVITICIGILMAPNVEVNSDFA  
 139, .1578

/product="nicotinic acetylcholine receptor alpha7 subunit"

3'UTR 1582, .1876 531 g 423 t  
 BASE COUNT 369 a 533 c 531 g 423 t  
 ORIGIN

Query Match 83.4%; Score 83.4; DB 10; Length 1876;  
 Best Local Similarity 98.8%; Pred. No. 1,6e-20;  
 Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 AGTGTGATGAGCGCTTGACGCACATTCACACTAGCTGTGTGATTTCTTGGG 68  
 421 AGTGTGATGAGCGCTTGACGCACATTCACACTAGCTGTGTGATTTCTTGGG 480  
 VY 69 CATGGCAGTACTGCTCCAGGTA 93  
 DB 481 CATGGCAGTACTGCTCCAGGCA 505

RESULT 6  
 LOCUS AF029838 641 bp mRNA PRI 16-DEC-1998  
 DEFINITION Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.  
 ACCESSION AF029838  
 NID 93757793  
 VERSION AF029838.1 GI:3757793  
 KEYWORDS  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 641)  
 AUTHORS Gault,J., Robinson,M., Berger,R., Drebing,C., Logel,J., Hopkins,J.,  
 Moore,T., Jacobs,S., Meriwether,J., Choi,M.J., Kim,E.J., Walton,K.,  
 Bulding,K., Davis,A., Breese,C., Freedman,R. and Leonard,S.  
 Genomic organization and partial duplication of the human alpha7  
 neuronal nicotinic acetylcholine receptor gene

TITLE JOURNAL  
 MEDLINE Genomics 52 (2), 173-185 (1998)  
 REFERENCE 99000837  
 JTHORS 2 (bases 1 to 641)  
 Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R.,  
 Breese,C., Davis,A., Hopkins,J. and Freedman,R.  
 Direct Submission  
 Submitted (15-OCT-1997) Psychiatry, University of Colorado Health  
 Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA

FEATURES  
 SOURCE  
 1. 641  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="15"  
 /map="15q14"  
 /clone="CHRNA7-DR1"  
 /note="RACE product A/B/C/D"  
 <1. >641  
 /note="alpha 7 neuronal nicotinic receptor mRNA; used in  
 alternative splicing"

## misc\_feature

BASE COUNT 151 a 192 c 163 g 135 t  
 ORIGIN

Query Match 82.8%; Score 82.8; DB 11; Length 641;  
 Best Local Similarity 97.7%; Pred. No. 2.4e-20;  
 Matches 84; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

3ATGAGCGCTTGAGCGCACATTCACACACAGCTGTGTGATTTCTTCTGG 67  
 "GAGCGCTTGAGCGCACATTCACACACAGCTGTGTGATTTCTTCTGG 591

QY 68 GCATTGCCAGTACTGCTCCAGGTA 93  
 DB 592 GCATTGCCAGTACTGCTCCAGGCA 617

RESULT 7  
 LOCUS AF029839 689 bp mRNA PRI 16-DEC-1998  
 DEFINITION Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.  
 ACCESSION AF029839  
 NID 93757794  
 VERSION AF029839.1 GI:3757794  
 KEYWORDS  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 689)  
 AUTHORS Gault,J., Robinson,M., Berger,R., Drebing,C., Logel,J., Hopkins,J.,  
 Moore,T., Jacobs,S., Meriwether,J., Choi,M.J., Kim,E.J., Walton,K.,  
 Bulding,K., Davis,A., Breese,C., Freedman,R. and Leonard,S.  
 Genomic organization and partial duplication of the human alpha7  
 neuronal nicotinic acetylcholine receptor gene

TITLE JOURNAL  
 MEDLINE Genomics 52 (2), 173-185 (1998)  
 REFERENCE 99000837  
 JTHORS 2 (bases 1 to 689)  
 Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R.,  
 Breese,C., Davis,A., Hopkins,J. and Freedman,R.  
 Direct Submission  
 Submitted (15-OCT-1997) Psychiatry, University of Colorado Health  
 Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA

FEATURES  
 SOURCE  
 1. 689  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="15"  
 /map="15q14"  
 /clone="CHRNA7-DR2"  
 /note="RACE product A/C/D"  
 <1. >689  
 /note="alpha 7 neuronal nicotinic receptor mRNA; used in  
 alternative splicing"

BASE COUNT 142 a 210 c 196 g 141 t  
 ORIGIN

Query Match 82.8%; Score 82.8; DB 11; Length 689;  
 Best Local Similarity 97.7%; Pred. No. 2.4e-20;  
 Matches 84; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 TAGTGTGATGAGCGCTTGAGCGCACATTCACACTAGCTGTGTGATTTCTTCTGG 67  
 DB 580 TAGTGTGATGAGCGCTTGAGCGCACATTCACACTAGCTGTGTGATTTCTTCTGG 639  
 QY 68 GCATTGCCAGTACTGCTCCAGGTA 93  
 DB 640 GCATTGCCAGTACTGCTCCAGGCA 665

RESULT 8  
 LOCUS AF036903 1712 bp mRNA PRI 16-OCT-1998  
 DEFINITION Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor  
 mRNA, alternatively spliced, partial sequence.  
 ACCESSION AF036903  
 NID 93757807  
 VERSION AF036903.1 GI:3757807  
 KEYWORDS  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1712)

AUTHORS Gault, J., Robinson, M., Berger, R., Drebling, C., Logel, J., Hopkins, J., Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K., Bulting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.  
TITLE Genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene (CHRNA7)  
JOURNAL Genomics 52 (2), 173-185 (1998)  
MEDLINE 99000837  
REFERENCE 2 (bases 1 to 1712)  
AUTHORS Leonard, S., Gault, J., Logel, J., Drebling, C., Robinson, M., Berger, R., Breese, C., Davis, A., Hopkins, J. and Freedman, R.  
TITLE Direct Submission  
JOURNAL Submitted (03-DEC-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA  
FEATURES  
SOURCE Location/Qualifiers  
1..1712  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/db\_xref="dbEST:W52861"  
/db\_xref="dbEST:AA037389"  
/chromosome="15"  
/clone\_11b="Soares papathyroid tumor NBHPA library, Research Genetics/IMAGE Consortium, LML"  
/clone\_11b="Soares papathyroid tumor NBHPA library, Research Genetics/IMAGE Consortium, LML"  
1..1712  
/note="alpha-7 neuronal nicotinic acetylcholine receptor; alternatively spliced"  
BASE COUNT 356 a 489 c 457 g 410 t  
ORIGIN

Query Match 82.8%; Score 82.8; DB 11; Length 1712;  
Best Local Similarity 97.7%; Pred. No. 2.6e-20;  
Matches 84; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 8 TAGTGTGATGAGCGCTTTGACGCCACATTCACACTAAGCTGTGGATCTCTGG 67  
|||  
DB 92 TATTGCTGATGAGCGCTTTGACGCCACATTCACACTAAGCTGTGGATCTCTGG 151  
|||  
OY 68 GCATTGCCAGTACCTGCTCCAGGTA 93  
|||  
DB 152 GCATTGCCAGTACCTGCTCCAGGTA 177  
|||

RESULT 9  
AF037646 1836 bp RNA PRI 16-OCT-1998  
LOCUS Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor  
DEFINITION precursor RNA, partial sequence.  
VERSION AF037646  
KEYWORDS AF037646.1 GI:3757808  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 1836)  
AUTHORS Gault, J., Robinson, M., Berger, R., Drebling, C., Logel, J., Hopkins, J., Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K., Bulting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.  
TITLE Genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene (CHRNA7)  
JOURNAL Genomics 52 (2), 173-185 (1998)  
MEDLINE 99000837  
REFERENCE 2 (bases 1 to 1836)  
AUTHORS Leonard, S., Gault, J., Logel, J., Drebling, C., Robinson, M., Berger, R., Breese, C., Davis, A., Hopkins, J. and Freedman, R.  
TITLE Direct Submission  
JOURNAL Submitted (03-DEC-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA  
FEATURES  
SOURCE Location/Qualifiers  
1..1836  
/organism="Homo sapiens"

misc-feature  
BASE COUNT 383 a 503 c 469 g 481 t  
ORIGIN

Query Match 82.8%; Score 82.8; DB 11; Length 1836;  
Best Local Similarity 97.7%; Pred. No. 2.6e-20;  
Matches 84; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 8 TAGTGTGATGAGCGCTTTGACGCCACATTCACACTAAGCTGTGGATCTCTGG 67  
|||  
DB 214 TATTGCTGATGAGCGCTTTGACGCCACATTCACACTAAGCTGTGGATCTCTGG 273  
|||  
OY 68 GCATTGCCAGTACCTGCTCCAGGTA 93  
|||  
DB 274 GCATTGCCAGTACCTGCTCCAGGTA 299  
|||

RESULT 10  
HSU40583 1977 bp mRNA PRI 19-DEC-1995  
LOCUS Human alpha 7 neuronal nicotinic acetylcholine receptor mRNA,  
DEFINITION complete cds.  
ACCESSION U40583  
NID 91125076  
VERSION U40583.1 GI:1125076  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 1977)  
AUTHORS Logel, J., Drebling, C., Barnhart, M., Antle, C. and Leonard, S.  
TITLE Nucleotide Sequence and Transcript Size of the Alpha-7 Neuronal Nicotinic Acetylcholine Receptor in Human Postmortem Brain  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1977)  
AUTHORS Leonard, S.  
TITLE Direct Submission  
JOURNAL Submitted (13-NOV-1995) Sherry Leonard, University of Colorado Health Sciences Center, C-268-71 Pharmacology, 4200 E. Ninth Ave., Denver, CO 80262, USA  
FEATURES  
SOURCE Location/Qualifiers  
1..1977  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="pshalpa7"  
/clone\_11b="Clontech lambda gt10 cDNA library: human hippocampus"  
/sex="male"  
/tissue-type="hippocampus"  
/dev-stage="20 yr old adult"  
1..7  
8..1516  
/codon\_start=1  
/product="alpha 7 neuronal nicotinic acetylcholine receptor"  
/protein\_id="AA83561.1"  
/db\_xref="PID:g1125077"  
/db\_xref="GI:1125077"  
/translation="MRCSPGVWLAALASLHVSLOGEPORKLYKELVKNYNPLEPY  
ANDSQPLTVFSLNLQIMDYDEKNOVLTININILQMSKWDHYLQMNVSERPGVITYRE

PDGQIMKPDILLVNSADERPDATFHTNVLNPSGHCQYLPGIKRSEKCYIDVREPPD  
 VOHCKLFGSMYSYGSMSLDLMOEADISGYPINPENGMDVIGIPGRSEFTFECCEPFB  
 DVTETVNRRTTYLNGLLIPCVLISLALVLELPDSEKISLGTIVLLSLTVM  
 LVAEIMPATSDSPVPLIAQVASTMIIVGLSVVTVIVLYQHNDHDPDGKMPKTRVY  
 LLNWCAMFLMRKRGEDVRPACQHKORCSLASVEKSAVPPASNGNLLYIGFRL  
 DGVHCVPTPDGSGVRMACSPTEHDEHLMHGSGPEGPDPAKILEEVRVIANFRCO  
 DESEVISEMKRACVDRCLMAFSVTIICITIGILMSAPNFEAVSKDPA"

3'UTR  
 polyA\_site 1514..1977  
 BASE COUNT 426 a 567 c 524 g 460 t  
 ORIGIN

Query Match 81.8% Score 81.8: DB 10: Length 1977;  
 Best Local Similarity 97.6% Pred. No. 6.2e-20;  
 Matches 83: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

9 AGTGTGATGAGCGCTTGACGCCACATTCACACTAGCTTGATGATCTCTGGG 68  
 356 AGTGTGATGAGCGCTTGACGCCACATTCACACTAGCTTGATGATCTCTGGG 415

QY 69 CATGCCAGTACCTGCTCCAGGTA 93  
 DB 416 CATGCCAGTACCTGCTCCAGGTA 440

RESULT 11  
 MUSNARS 1848 bp mRNA ROD 02-FEB-1999  
 LOCUS Mus musculus neuronal nicotinic acetylcholine receptor subunit  
 DEFINITION alpha 7 mRNA, complete cds.  
 ACCESSION L37663  
 NID L37663.1 GI:790853  
 VERSION L37663.1  
 KEYWORDS neuronal acetylcholine receptor subunit alpha 7 unit;  
 SOURCE Mus musculus (strain BALB/c, sub-species domesticus) neonatal brain  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 1848)  
 AUTHORS Orr-Urtreger, A., Seldin, M.F., Baldini, A. and Beaudet, A.L.  
 TITLE Cloning and mapping of the mouse alpha 7-neuronal nicotinic  
 acetylcholine receptor  
 JOURNAL Genomics 26 (2), 399-402 (1995)  
 MEDLINE 95324936

JOURNAL MEDLINE 95324936  
 TITLE Cloning and mapping of the mouse alpha 7-neuronal nicotinic  
 acetylcholine receptor  
 JOURNAL Genomics 26 (2), 399-402 (1995)  
 MEDLINE 95324936

location/Qualifiers  
 1. 1848  
 /organism="Mus musculus"  
 /strain="BALB/c"  
 /sub-species="domesticus"  
 /db\_xref="taxon:10090"  
 /dev\_stage="neonatal"  
 /tissue\_type="brain"  
 1..50  
 51..1559  
 /codon\_start=1  
 /product="neuronal acetylcholine receptor subunit alpha 7"  
 /protein\_id="AAC42053.1"  
 /db\_xref="PID:g790854"  
 /db\_xref="GI:790854"  
 /translation="MCGRRGIMLAAALHVSLOGEFORRLKELVKNPNLEPVP  
 ANDSPLTYFSLQIDMDVDEKQVLTINWLMQSMTHDYLQNMNSEYGVKNVRF  
 PPGQIKRPDILLVNSADERPDATFHTNVLNPSGHCQYLPGIKRSEKCYIDVREPPD  
 VOOCILKRGSMYSYGSMSLDLMOEADISGYPINPENGMDVIGIPGRSEFTFECCEPFB  
 DVTETVNRRTTYLNGLLIPCVLISLALVLELPDSEKISLGTIVLLSLTVM  
 LVAEIMPATSDSPVPLIAQVASTMIIVGLSVVTVIVLYQHNDHDPDGKMPKTRVY  
 LLNWCAMFLMRKRGEDVRPACQHKORCSLASVEKSAVPPASNGNLLYIGFRL  
 DGVHCVPTPDGSGVRMACSPTEHDEHLMHGSGPEGPDPAKILEEVRVIANFRCO  
 DESEVISEMKRACVDRCLMAFSVTIICITIGILMSAPNFEAVSKDPA"

5'UTR  
 CDS  
 3'UTR  
 BASE COUNT 406 a 487 c 500 g 455 t  
 ORIGIN

Query Match 69.0% Score 69: DB 12: Length 1848;  
 Best Local Similarity 88.2% Pred. No. 3.1e-15;  
 Matches 75: Conservative 0; Mismatches 10; Indels 0; Gaps 0;

9 AGTGTGATGAGCGCTTGACGCCACATTCACACTAGCTTGATGATCTCTGGG 68  
 399 AGTGTGATGAGCGCTTGACGCCACATTCACACTAGCTTGATGATCTCTGGG 458

QY 69 CATGCCAGTACCTGCTCCAGGTA 93  
 DB 459 CATGCCAGTACCTGCTCCAGGTA 483

RESULT 12  
 RATNARAD 2106 bp mRNA ROD 28-DEC-1998  
 LOCUS Rattus rattus nicotinic acetylcholine receptor alpha 7 subunit  
 DEFINITION mRNA, complete cds.  
 ACCESSION L31619  
 NID L31619.1 GI:3478618  
 VERSION L31619.1  
 KEYWORDS Norway rat.  
 SOURCE Rattus norvegicus  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 REFERENCE 1 (bases 1 to 2106)  
 AUTHORS Boulter, J.  
 TITLE Nicotinic acetylcholine receptor subunit gene alpha 7  
 JOURNAL Unpublished (1994)  
 REFERENCE 2 (bases 1 to 2106)  
 AUTHORS Boulter, J.  
 JOURNAL Unpublished (1992)  
 REFERENCE 3 (bases 1 to 2106)  
 AUTHORS Boulter, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-APR-1994) Department of Psychiatry and Biobehavioral  
 Sciences, University of California, 760 Westwood Plaza, Los  
 Angeles, CA 90095-1759, USA  
 4 (bases 1 to 2106)  
 REFERENCE 4 (bases 1 to 2106)  
 AUTHORS Hartley, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-AUG-1998) The Salk Institute, MNJ-H, 10010 North  
 Torrey Pines Road, La Jolla, CA 92037, USA  
 COMMENT Sequence update by submitter  
 ON Aug 27, 1998 this sequence version replaced gi:468919.  
 FEATURES  
 location/Qualifiers  
 1..2106  
 /organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /cell\_type="neuronal"  
 /dev\_stage="adult"  
 /tissue\_type="brain"  
 23..1531  
 /codon\_start=1  
 /product="nicotinic acetylcholine receptor alpha 7  
 subunit"  
 /protein\_id="AAC33136.1"  
 /db\_xref="PID:g3478619"  
 /db\_xref="GI:3478619"  
 /translation="MCGRRGIMLAAALHVSLOGEFORRLKELVKNPNLEPVP  
 ANDSPLTYFSLQIDMDVDEKQVLTINWLMQSMTHDYLQNMNSEYGVKNVRF  
 PPGQIKRPDILLVNSADERPDATFHTNVLNPSGHCQYLPGIKRSEKCYIDVREPPD  
 VOOCILKRGSMYSYGSMSLDLMOEADISGYPINPENGMDVIGIPGRSEFTFECCEPFB  
 DVTETVNRRTTYLNGLLIPCVLISLALVLELPDSEKISLGTIVLLSLTVM  
 LVAEIMPATSDSPVPLIAQVASTMIIVGLSVVTVIVLYQHNDHDPDGKMPKTRVY  
 LLNWCAMFLMRKRGEDVRPACQHKORCSLASVEKSAVPPASNGNLLYIGFRL  
 DGVHCVPTPDGSGVRMACSPTEHDEHLMHGSGPEGPDPAKILEEVRVIANFRCO  
 DESEVISEMKRACVDRCLMAFSVTIICITIGILMSAPNFEAVSKDPA"

BASE COUNT 473 a 554 c 533 g 546 t

## ORIGIN

Query Match 69.0%; Score 69; DB 12; Length 2106;  
 Best Local Similarity 88.2%; Pred. No. 3.2e-15;  
 Matches 75; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 9 AGTCGTATGAGCGCTTGACGACATTCACACTAAGCTGGTGAATCTCTGGG 68  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 371 AGTCGTATGAGCGCTTGACGACATTCACACTAAGCTGGTGAATCTCTGGG 430  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 69 CATGCCAGTACCTGCTCCAGGTA 93  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 431 CATGCCAGTACCTGCTCCAGGTA 455

## RESULT 13

LOCUS S53987 3030 bp mRNA ROD 19-MAR-1993  
 INITIATION nicotinic receptor alpha 7 subunit [rats, brain, mRNA, 3030 nt].  
 ESSION S53987  
 NID Q264770  
 VERSION S53987.1 GI:264770  
 KEYWORDS  
 SOURCE Rattus sp. brain.  
 ORGANISM Rattus sp.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

## REFERENCE

AUTHORS Seguela, P., Wadiche, J., Dineley-Miller, K., Dani, J.A. and Patrick, J.W.  
 TITLE Molecular cloning, functional properties, and distribution of rat brain alpha 7: a nicotinic cation channel highly permeable to calcium  
 J. Neurosci. 13 (2), 596-604 (1993)

## JOURNAL

MEDLINE 93147931  
 REMARK GenBank staff at the National Library of Medicine created this entry [NCBI g1dbseq 124020] from the original journal article.

## FEATURES

source  
 1. 3030  
 location/Qualifiers  
 /organism="Rattus sp."  
 /db\_xref="taxon:10118"  
 42. 1550  
 /gene="nicotinic receptor alpha 7 subunit"  
 42. 1550  
 /note="This sequence comes from Fig. 1. Protein sequence is in conflict with the conceptual translation;  
 mismatch(469[R->P])"  
 /codon\_start=1  
 /product="nicotinic receptor alpha 7 subunit"  
 /protein\_id="AA82524.1"  
 /db\_xref="PID:g264771"  
 /db\_xref="GI:264771"  
 /translation="MCGRGIGWLAALAAALLHVSLOGEFORRLYKYLKYNPLERPV  
 ANDSPITVYFSLQIDVDENQVLTNTIMLQMSWDHYLQNNSEYGVKNVFE  
 PDSOIMKPDILLYNSADDERFATFTNVNLSGHCQYLPGLFKSSCYIDVMPEDV  
 VOOCKIKKSSYGGSLDLOMOEADISYINGEMDLVGVKSKSEKYECKEPEPV  
 DVYVYTRRRRLTYGNNLITPCVLSALALVFLPADSGKISGLITVLSIVFMIL  
 LVAELMPATSDVPLIAQYFASMTIVGLSVVTVIVLQHHDPDGGKMKWTRVIL  
 LKWCAMFLMKRKGEDKVRPAQOHKRRCSLASVEMSAVAPPTNGNLVIGRGLDTM  
 EGHNCAPTSDSGVCGRLASPTDEHLMHAGPSGPDLAKEVRYIAHRPDCODES  
 DESEVISEKFAVCYDRLCLMAFSVFTICTIGILMSAPNVEAVSKDA"

## CDS

gene  
 42. 1550  
 /gene="nicotinic receptor alpha 7 subunit"  
 /note="This sequence comes from Fig. 1. Protein sequence is in conflict with the conceptual translation;  
 mismatch(469[R->P])"  
 /codon\_start=1  
 /product="nicotinic receptor alpha 7 subunit"  
 /protein\_id="AA82524.1"  
 /db\_xref="PID:g264771"  
 /db\_xref="GI:264771"  
 /translation="MCGRGIGWLAALAAALLHVSLOGEFORRLYKYLKYNPLERPV  
 ANDSPITVYFSLQIDVDENQVLTNTIMLQMSWDHYLQNNSEYGVKNVFE  
 PDSOIMKPDILLYNSADDERFATFTNVNLSGHCQYLPGLFKSSCYIDVMPEDV  
 VOOCKIKKSSYGGSLDLOMOEADISYINGEMDLVGVKSKSEKYECKEPEPV  
 DVYVYTRRRRLTYGNNLITPCVLSALALVFLPADSGKISGLITVLSIVFMIL  
 LVAELMPATSDVPLIAQYFASMTIVGLSVVTVIVLQHHDPDGGKMKWTRVIL  
 LKWCAMFLMKRKGEDKVRPAQOHKRRCSLASVEMSAVAPPTNGNLVIGRGLDTM  
 EGHNCAPTSDSGVCGRLASPTDEHLMHAGPSGPDLAKEVRYIAHRPDCODES  
 DESEVISEKFAVCYDRLCLMAFSVFTICTIGILMSAPNVEAVSKDA"

## BASE COUNT

698 a 756 c 801 g 775 t

Query Match 69.0%; Score 69; DB 12; Length 3030;  
 Best Local Similarity 88.2%; Pred. No. 3.2e-15;  
 Matches 75; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 9 AGTCGTATGAGCGCTTGACGACATTCACACTAAGCTGGTGAATCTCTGGG 68

DB 390 AGTCGTATGAGCGCTTGACGACATTCACACTAAGCTGGTGAATCTCTGGG 449  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 69 CATGCCAGTACCTGCTCCAGGTA 93  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 450 CATGCCAGTACCTGCTCCAGGTA 474

## RESULT 14

LOCUS BTA7 1551 bp mRNA NAM 09-JAN-1996  
 INITIATION B. taurus mRNA for alpha7 nicotinic acetylcholine receptor subunit.  
 ACCESSION X93604  
 NID 91181202  
 VERSION X93604.1 GI:1181202  
 KEYWORDS alpha7 gene; alpha7 nicotinic receptor; alpha7 subunit; nicotinic acetylcholine receptor.  
 SOURCE Bos taurus.  
 ORGANISM Bos taurus.  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Artiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

## REFERENCE

AUTHORS Garcia-Guzman, M., Sala, F., Sala, S., Campos-Caro, A., Stuhmer, W., Gutierrez, L.M. and Criado, M.  
 TITLE alpha-Bungarotoxin-sensitive nicotinic receptors on bovine chromaffin cells: molecular cloning, functional expression and alternative splicing of the alpha 7 subunit  
 Eur. J. Neurosci. 7 (4), 647-655 (1995)

## JOURNAL

MEDLINE 9534609  
 REMARK 2 (bases 1 to 1551)  
 AUTHORS Garcia-Guzman, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-NOV-1995) M. Garcia-Guzman, Max-Planck Institut fuer Experimentelle Medizin (Abt. XI), Hermann-Rein Strasse, 3, D-37075-Goettingen, FRG  
 REVISED by [3]  
 On Feb 6, 1996 this sequence version replaced g1:1103382.

## FEATURES

source  
 1. 1551  
 location/Qualifiers  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /dev\_stage="adult"  
 /tissue\_type="adrenal gland medulla"  
 43. 99  
 /gene="alpha7"  
 43. 1542  
 /gene="alpha7"  
 /codon\_start=1  
 /product="alpha7 nicotinic acetylcholine receptor subunit"  
 /protein\_id="CAA63802.1"  
 /db\_xref="PID:g212821"  
 /db\_xref="PID:g1103383"  
 /db\_xref="GI:1103383"  
 /db\_xref="SWISS-PROT:P54131"  
 /translation="MRGSLCLAAASILHVSLOGEFORRLYKYLKYNPLERPVAND  
 SLPITVYFSLQIDVDENQVLTNTIMLQMSWDHYLQNNSEYGVKNVFE  
 PDSOIMKPDILLYNSADDERFATFTNVNLSGHCQYLPGLFKSSCYIDVMPEDV  
 VOOCKIKKSSYGGSLDLOMOEADISYINGEMDLVGVKSKSEKYECKEPEPV  
 DVYVYTRRRRLTYGNNLITPCVLSALALVFLPADSGKISGLITVLSIVFMIL  
 LVAELMPATSDVPLIAQYFASMTIVGLSVVTVIVLQHHDPDGGKMKWTRVIL  
 LKWCAMFLMKRKGEDKVRPAQOHKRRCSLASVEMSAVAPPTNGNLVIGRGLDTM  
 EGHNCAPTSDSGVCGRLASPTDEHLMHAGPSGPDLAKEVRYIAHRPDCODES  
 DESEVISEKFAVCYDRLCLMAFSVFTICTIGILMSAPNVEAVSKDA"

## CDS

sig-peptide  
 43. 99  
 /gene="alpha7"  
 43. 1542  
 /gene="alpha7"  
 /codon\_start=1  
 /product="alpha7 nicotinic acetylcholine receptor subunit"  
 /protein\_id="CAA63802.1"  
 /db\_xref="PID:g212821"  
 /db\_xref="PID:g1103383"  
 /db\_xref="GI:1103383"  
 /db\_xref="SWISS-PROT:P54131"  
 /translation="MRGSLCLAAASILHVSLOGEFORRLYKYLKYNPLERPVAND  
 SLPITVYFSLQIDVDENQVLTNTIMLQMSWDHYLQNNSEYGVKNVFE  
 PDSOIMKPDILLYNSADDERFATFTNVNLSGHCQYLPGLFKSSCYIDVMPEDV  
 VOOCKIKKSSYGGSLDLOMOEADISYINGEMDLVGVKSKSEKYECKEPEPV  
 DVYVYTRRRRLTYGNNLITPCVLSALALVFLPADSGKISGLITVLSIVFMIL  
 LVAELMPATSDVPLIAQYFASMTIVGLSVVTVIVLQHHDPDGGKMKWTRVIL  
 LKWCAMFLMKRKGEDKVRPAQOHKRRCSLASVEMSAVAPPTNGNLVIGRGLDTM  
 EGHNCAPTSDSGVCGRLASPTDEHLMHAGPSGPDLAKEVRYIAHRPDCODES  
 DESEVISEKFAVCYDRLCLMAFSVFTICTIGILMSAPNVEAVSKDA"

## BASE COUNT

305 a 477 c 429 g 340 t

Query Match 67.4%; Score 67.4; DB 3; Length 1551;  
 Best Local Similarity 87.1%; Pred. No. 1.2e-14;

QY 9 AGTCGTATGAGCGCTTGACGACATTCACACTAAGCTGGTGAATCTCTGGG 68

Matches 74; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 9 AGTGTGATGAGCGCTTTGAGCGCCATTCACACTACAGTGTGTAATCTCTGGG 68  
|||||  
Db 352 AGTGTGATGAGCGCTTTGAGCGCCATTCACACTACAGTGTGTAATCTCTGGG 441

QY 69 CATGCCAGTACCTGCTCCAGGTA 93  
|||  
Db 442 CACTGCCAGTACCTGCTCCAGGTA 466

## RESULT 15

GGA7NARC 2696 bp mRNA VRT 12-AUG-1996

LOCUS Chicken alpha7 subunit of nicotinic acetylcholine receptor.

ACCESSION X52295

NID 963077

VERSION X52295.1 GI:63077

WORDS alpha7 subunit; nicotinic acetylcholine receptor.

ORIGIN chicken.

ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;

Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 2696)

AUTHORS Schoepfer, R.

TITLE Direct Submission

JOURNAL Submitted (12-APR-1990) Schoepfer R., The Salk Institute for

Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,

USA

2 (bases 1 to 2696)

AUTHORS Schoepfer, R., Conroy, W.G., Whitting, P., Gore, M., and Lindstrom, J.

TITLE Brain alpha-bungarotoxin binding protein cDNAs and Mabs reveal

subtypes of this branch of the ligand-gated ion channel gene

superfamily

JOURNAL Neuron 5 (1), 35-48 (1990)

MEDLINE 90315158

COMMENT See also <X52296>.

FEATURES Location/Qualifiers

1..2696

/organism="Gallus gallus"

/db\_xref="taxon:9031"

/dev\_stage="18 day Old embryo"

/tissue\_type="brain"

/clone\_id="pch29-3 and pch34-1."

5..70

/product="signal peptide (AA-22 to -1)"

5..1513

/note="protein precursor (AA-21-480)"

/codon\_start=1

/product="alpha7 subunit of nicotinic acetylcholine

receptor

/protein\_id="CAA36543.1"

/db\_xref="PIR:G63078"

/db\_xref="GI:63078"

/translation="MGLRALMLLAAAGLVRESLOGEPRKRYKELKYNLERPV

ANDSQPLVFTLSIMQINDYDERNOYLTINILQMTWIDHILQNNVSEPGVKNRF

PDGLINKFDLILNSADERFDATFHINVLVNSGHCQYLPPIGKSSCYIDVWPEPD

VQRCNLKFGSWTGLQMOEADISGINSQMDLVGIPEGKRESFECCKEP

DIFETVYRRRTLYGLNLPCVLISALALVLELPADSGEKISGITVLSLTVEM

LVAEIMPATSDVPLIAOYFASMTIIVGLSVVTVIYLOHHDPDGGKMPKTRVI

LNNKCAFLMRKRGEDKVRACQKORCSLSMEMNTYSGQCSNGNNLYIGFRGL

DGVHCPRTDSGVICGRTKSPTEERELHSGHPSGSDPDLAKILEVERIARFRDQ

Matches 68; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 9 AGTGTGATGAGCGCTTTGAGCGCCATTCACACTACAGTGTGTAATCTCTGGG 68  
|||||  
Db 353 AGTGTGATGAGCGCTTTGAGCGCCATTCACACTACAGTGTGTAATCTCTGGG 412

QY 69 CATGCCAGTACCTGCTCCAGGTA 93  
|||  
Db 413 CACTGCCAGTACCTGCTCCAGGTA 437

Search completed: September 17, 1999, 22:04:21  
Job time: 16444 sec

Query Match 57.8%; Score 57.8; DB 4; Length 2696;  
Best Local Similarity 80.0%; Pred. No. 4.3e-11;

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OK nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:34:33 ; Search time 425.19 seconds

(without alignments)  
58.842 Million cell updates/sec

Title: US-08-956-518A-99

Perfect score: 100

Sequence: 1 CTGTTCTATGCTGATGAG.....CTGCCCTCCAGGTAAGTCA 100

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database: N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83.4	83.4	1876	1 T48239	Neuronal nicotinic
2	83.4	83.4	1876	1 V12197	Human neuronal nic
3	83.4	83.4	1590	1 V44687	V2747 variant huma
4	57.8	57.8	2769	1 T59196	Neuronal alpha-bun
5	48.8	48.8	2101	1 T59197	Neuronal alpha-bun
6	30.6	30.6	2460	1 Q06086	Plasmid p2PC13 enc
7	28.4	28.4	2192	1 X23530	Matz x21 gene DT
8	26.6	26.6	1365	1 T79976	Human galanin rece
9	26.6	26.6	1083	1 V10329	Human GalR2 recept
10	26.6	26.6	3390	1 V44931	Human galanin rece
11	26.6	26.6	3390	1 V28291	Human galanin rece
12	26.6	26.6	3390	1 V22649	Human galanin rece
13	25.2	25.2	13574	1 X13051	Enterococcus faeca
14	24.8	24.8	5002	1 X25011	Anter-specific ES
15	24.4	24.4	2376	1 X16153	Mouse Sox1 cDNA. I
16	24.4	24.4	443	1 X21047	Polynucleotide seq
17	24.2	24.2	1934	1 Q91198	HMG 46 kDa antige
18	24.2	24.2	1384	1 Q91199	HMG 46 kDa antige
19	24.0	24.0	1734	1 T61134	Batten disease gen
20	24.0	24.0	1732	1 T61135	Batten disease gen
21	24.0	24.0	1732	1 T61136	Batten disease gen
22	24.0	24.0	1732	1 T61137	Batten disease gen
23	24.0	24.0	1732	1 T61138	Batten disease gen
24	24.0	24.0	1732	1 T61139	Batten disease gen
25	24.0	24.0	1732	1 T61140	Batten disease gen
26	24.0	24.0	1732	1 T61141	Batten disease gen
27	24.0	24.0	1732	1 T61142	Batten disease gen
28	24.0	24.0	1732	1 T61143	Batten disease gen
29	24.0	24.0	1732	1 T61144	Batten disease gen
30	24.0	24.0	1732	1 T61145	Batten disease gen
31	24.0	24.0	1732	1 T61146	Batten disease gen
32	24.0	24.0	1732	1 T61147	Batten disease gen
33	24.0	24.0	1732	1 T61148	Batten disease gen
34	24.0	24.0	1732	1 T61149	Batten disease gen
35	24.0	24.0	1732	1 T61150	Batten disease gen
36	24.0	24.0	1732	1 T61151	Batten disease gen
37	24.0	24.0	1732	1 T61152	Batten disease gen
38	23.8	23.8	2038	1 Q89844	Human death associ
39	23.6	23.6	11601	1 V60292	DNA sequence encod
40	23.6	23.6	13058	1 Q13608	ACV synthetase gen
41	23.6	23.6	13058	1 Q48231	Vector containing
42	23.6	23.6	13058	1 V52256	Streptococcus pneu
43	23.6	23.6	1569	1 V59650	Human secreted pro

44 23.4 23.4 3039 1 Q91096 Human herpesvirus-  
c 45 23.4 23.4 8738 1 T72327 Lactobacillus bact

## ALIGNMENTS

RESULT 1  
ID T48239 standard; DNA; 1876 BP.  
AC T48239;  
DT 09-APR-1997 (first entry)  
DE Neuronal nicotinic acetylcholine receptor alpha-7 subunit DNA.  
KW Neuronal nicotinic acetylcholine receptor; nAChR; neurotransmitter;  
RW ligand-gated receptor; ds.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT cds 73..1581  
FT /\*tag= a  
FN W09641876-A1.  
PD 27-DEC-1996.  
PE 07-JUN-1996; U09775.  
PR 07-JUN-1995; US-484722.  
PA (SIBT-) SIBT NEUROSCIENCES INC.  
PI Elliott KJ, Harpold WM;  
DR P-PSDB; W09025.  
PT Nucleic acids encoding nicotinic acetylcholine receptor sub-units -  
used in screening to determine the effect of drugs on the receptor  
PS Disclosure; Page 71-73; 108pp; English.  
CC A DNA sequence (T48239) codes for the alpha-7 subunit (W09025) of  
the human neuronal nicotinic acetylcholine receptor (nAChR). Host  
cells, esp. mammalian cells or amphibian oocytes, carrying alpha-7  
nucleic acids, opt. in combination with other alpha and/or beta  
subunit nucleic acids (see also T48232-38, T48240-41), express  
recombinant nAChR subunits useful for identifying cpds. that  
modulate the activity of human nAChRs.  
CC Sequence 1876 BP; 369 A; 553 C; 531 G; 423 T;  
SQ

Query Match 83.4%; Score 83.4; DB 1; Length 1876;  
Best Local Similarity 98.8%; Pred. No. 5.6e-22;  
Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 AGTGTGATGAGCGCTTGAGCCACATTCACACTAAGTGTGATCTTCGG 68  
DB 421 AGTGTGATGAGCGCTTGAGCCACATTCACACTAAGTGTGATCTTCGG 480  
OY 69 CATTCACAGTACCTGCTCCAGGTA 93  
DB 481 CATTCACAGTACCTGCTCCAGGCA 505

RESULT 2  
ID V12197 standard; cDNA; 1876 BP.  
AC V12197;  
DT 14-MAY-1998 (first entry)  
DE Human neuronal nicotinic acetylcholine receptor alpha-7 subunit cDNA.  
KW Human; neuronal nicotinic acetylcholine receptor; alpha-7 subunit;  
RW brain tissue; screening; NACNR; antibody; ds.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT cds 73..1581  
FT /\*tag= a  
FT /product= "neuronal nicotinic acetylcholine receptor  
alpha-7 subunit".  
FN W09420617-A2.  
PD 15-SEP-1994.  
PE 08-MAR-1994; U02447.  
PR 08-MAR-1993; US-028031.  
PA (SIBT-) SIBT INST BIOTECHNOLOGY IND ASSOC.  
PI (SIBT-) SIBT NEUROSCIENCES INC.

PI	Elliott GJ, Ellis SB, Harpold KM;
DR	WPI: 94-303024/37
P-PDB:	W44153.
PT	Human neuronal nicotinic acetylcholine receptor subunits and DNA -
CC	also transformed cells useful for screening cpds. which modulate
PT	activity of the receptor
PS	Claim 8; Page 78-79; 99pp; English.
CC	The present sequence encodes a human neuronal nicotinic acetylcholine
CC	receptor (nAChR) subunit. The cells expressing the alpha and/or beta
CC	nAChR subunits may be used in a method of screening compounds to
CC	identify any which modulate the activity of human neuronal nAChR.
CC	Subunit specific antibodies may be used to monitor the distribution
CC	and expression density of various subunits in normal vs diseased brain
CC	tissues. Testing of single receptor subunits or specific receptor
CC	subunit combinations with a variety of potential agonists or antagonists
CC	provides information with respect to the function and activity of the
CC	individual subunits and should lead to the identification and design of
CC	compounds that are capable of very specific interaction with one or
CC	more receptor subtypes. The resulting drugs should exhibit fewer
CC	unwanted side effects than drugs identified e.g. screening with cells
-	that express a variety of subtypes.
-	Sequence 1876 BP: 369 A; 553 C; 530 G; 423 T;
Query Match	Best Local Similarity 83.4%; Score 83.4; DB 1; Length 1876;
Matches	84; Conservative 0; Mismatches 1; Indels 0; Gaps
OY	9 AGTCTGATGAGCGCCTTGAGCGCACATTCCACACTACGTGTGGTAATTCTTTGGG 68
Db	421 AGTCCTATATAGGCCCTTTGAGCGCACATTTCCACACTAACGTGTGGTAATTCTTTGGG 480
OY	69 CATGCCAGTAGCTGCCTCCAGGTA 93
Db	481 CATGCCAGTAGCTGCCTCCAGCA 505
RESULT	3
V44687	
ID	V44687 standard; cDNA; 1590 BP.
AC	V44687;
DE	09-OCT-1998 (first entry)
DT	V274T variant human alpha7 nAChR coding sequence.
KM	Alpha7 nAChR: alpha7 nicotinic acetylcholine receptor subunit; cancer;
KW	neurodegeneration; enzyme dysfunction; affective disorder; therapy;
KW	immune dysfunction; diabetic neuropathy; Alzheimer's disease;
KW	sclerophrenia; ss.
Homo sapiens.	
Key	Location/Qualifiers
CDS	9..1517
FT	/tag= a
PN	MO92828331-A2.
PD	02-JUL-1998.
PF	22-DEC-1997; U23405.
PA	20-DEC-1996; US-771737.
PR	(ABBO ) ABBOTT LAB.
PI	Biggs CA, Gopalakrishnan M, McKenna DG, Monteggia LM,
J1	Roch J, Sullivan JP, Touma E;
WPI:	98-377593/32.
DR	P-PDB: W69216.
CC	Nucleic acid encoding variant of human alpha7 nicotinic
CC	acetyl-choline receptor sub-unit - used to identify modulators of
CC	the receptor, potentially useful for treating neuro-degeneration,
CC	cancer, affective disorders etc.
CC	Claim 14; Fig 2; 44pp; English.
CC	This sequence encodes the V247T variant of human alpha7 nicotinic
CC	acetylcholine receptor (nAChR) subunit of the invention. Cells containing
CC	the DNA are used to express the protein and to identify modulators of
CC	alpha nAChR activity or cytoprotective agents, e.g. antisense
CC	compounds or antagonists that are potentially useful for treating
CC	neurodegeneration, enzyme dysfunction, affective disorders and immune
CC	dysfunction, such as cancer, post-hepetic neuritis, diabetic
CC	neuropathy, osteoarthritis, Alzheimer's or Parkinson's diseases, kuru,

```

CC psychositis and schizophrenia. Probes based on the DNA are used to detect
CC the DNA in usual hybridisation or amplification tests, while monoclonal
CC antibodies are used to detect the protein for diagnosis (in vitro or by
CC in situ immuno-fluorescent assay). Compared with wild-type alpha7 nAChR,
CC the protein has about 100-fold greater sensitivity to cholinergic
CC receptor agonists (nicotine or acetylcholine) and response to these
CC agonists decays more slowly, but the wild-type inward rectification is
CC retained.
S0 Sequence 1590 BP; 322 A; 471 C; 449 G; 348 T;

Query Match 83.4%; Score 83.4; DB 1; Length 1590;
Best Local Similarity 98.8%; Pred.No.5,3e-22;
Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0

QY 9 AGTGGCTATAGCGCTTTGACGCCACATTCACACTAACGTTGTGTAATCTTCGGG 68
DB 357 AGTGGCTATAGCGCTTTGACGCCACATTCACACTAACGTTGTGTAATCTTCGGG 416
QY 69 CATGGCCAGTACCTGCCTCCAGGTA 93
DB 417 CATGGCCAGTACCTGCCTCCAGGCA 441

RESULT 4
ID T59196 standard; cDNA: 2769 BP.
AC T59196;
DT 17-JUN-1997 (first entry)
DE Neuronal alpha-bungarotoxin binding protein alpha1 subunit cDNA.
KW Neuronal alpha-bungarotoxin binding protein alpha 1; cholinergic;
OS Gallus sp.
FH Key Location/Qualifiers
FT cds 71..1513
FT signal_peptide 71..136
FT mat_peptide 137..1510
FT tag_c
PD US5599709-A.
PN 04-FEB-1997.
PR 28-SEP-1989; 413947.
PA (SALK) SALK INST BIOLOGICAL STUDIES.
PI Lindstrom JM, Schoepfer RD;
DR WPI; 97-118297/11.
DR P-PDB: W12368.
PT New isolated neuronal alpha-bungarotoxin-binding protein DNA - used
PT to screen cholinergic agents and other drugs which may affect ligand
PT binding, ion channel or other activities of the protein.
PS Claim 1; Fig 2A-B; 18pp; English.
CC 2 cDNA clones (T59196 and T59197) respectively code for the alpha1
CC subunit (W12368) and alpha2 subunit (W12369) of chick neuronal
CC alpha-bungarotoxin binding protein (ABBP). They were isolated from
CC an 18-day embryo chick brain cDNA library using a probe (see also
CC T59198) based on the N-terminal amino acid sequence of chicken
CC brain ABBP. The probe isolated partial clone pCh29-1, which
CC encoded the N-terminal portion of alpha1. A subclone, pCh29-3
CC (ATCC 40641), was used to rescreen the library, yielding clone
CC pCh31-1 (ATCC 40640), which encoded the entire alpha2 sequence.
CC A probe based on the C-terminal region of pCh31-1 was used obtain
CC clone pCh34-1 (ATCC 40639), encoding the C-terminal portion of
CC alpha1. The cDNA clones can be used as probes to identify further
CC ABBP subunits, and in the recombinant prodn. of ABBP.
S0 Sequence 2769 BP; 790 A; 604 C; 571 G; 804 T;

Query Match 57.8%; Score 57.8; DB 1; Length 2769;
Best Local Similarity 80.0%; Pred.No.2,4e-12;
Matches 68; Conservative 0; Mismatches 17; Indels 0; Gaps 0

Y 9 AGTGGCTATAGCGCTTTGACGCCACATTCACACTAACGTTGTGTAATCTTCGGG 68

```





RESULT 7  
X23530 standard; DNA; 2192 BP.

AC X23530;  
DE 17-JUN-1999 (first entry)  
KW Maize Xa21 gene D74 DNA fragment.  
KW Xa21: receptor kinase-like protein; multigene family; RRR; rice; D74;  
KW plant disease resistance; cassava; maize; tomato; Xanthomonas; ss.  
OS Zea mays.  
PN MO909151-A2.  
PD 25-FEB-1999.  
PF 17-JUL-1998; U14841.  
PR 13-AUG-1997; US-910386.  
PA (REGC) UNIV CALIFORNIA.  
PI Hubert SH, Richter T, Ronald PC, Song W, Szabo V,  
PI Wang G;  
PI WPI: 99-204431/17.  
PT New RRR polynucleotides and nucleic acid constructs - used for  
PT generating transgenic plants resistant to Xanthomonas  
PT Claim 12; Page 57-58; 67pp; English.  
CC This invention describes a method for conferring disease resistance in  
CC plants. The invention describes the use of novel genes and proteins  
CC belonging to the Oryza longistaminata and Oryza sativa receptor  
CC kinase-like protein (RRK) Xa21 multigene family. Such genes from  
CC cassava, maize and tomato are also described. The genes and proteins can  
CC be used for enhancing resistance to Xanthomonas in a plant, preferably in  
CC rice or tomato.  
SQ Sequence 2192 BP: 422 A; 659 C; 693 G; 418 T;

Query Match 28.4%; Score 28.4; DB 1; Length 2192;  
Best Local Similarity 58.1%; Pred. No. 0.23;  
Matches 50; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 13 CTGATGAGCGCTTTGACGCCACATTCACACTACGCTGTGGAATCTCTGGGCATT 72  
DB 645 CTGCTGGCGCGGCTGCCCTCCAGCTCGGATGAGTAGTGTGTCGACGTGACCTG 704  
OY 73 GCCAGTACCTGCTCCGACGATGACCTG 98  
DB 705 TCCAGCACCCAGCTGAGGAGGAGAGCTG 730

RESULT 8  
T79976/c  
ID T79976 standard; cDNA; 1365 BP.  
AC T79976;  
DE 04-FEB-1998 (first entry)  
KW Human galanin receptor GALT2 cDNA.  
KW Galanin receptor; GALT2; human; agonist; antagonist; pain;  
KW eating disorder; Alzheimer's disease; diabetes; ischemia;  
KW obesity; bulimia; anorexia; nociception; cognitive disorder;  
KW sensory disorder; motion sickness; hypertension; convulsion;  
KW epilepsy; glaucoma; ulcer; inflammation; immune disease; anxiety;  
KW therapy; ss.  
OS Homo sapiens.  
PN Key  
FT CDS Location/Qualifiers  
FT 102..1265  
FT /\*tag= a  
FT MO9726853-A2.  
PD 31-JUL-1997.  
PF 24-JAN-1997; U01301.  
PR 27-SEP-1996; US-721837.  
PR 24-JAN-1996; US-590494.  
PR 01-APR-1996; US-626046.  
PR 01-APR-1996; US-626885.  
PA (SYNA-) SYNAPTIC PHARM CORP.  
PI Branchek T, Forray C, Gerald CPG, Linemeyer D, Smith KE;  
PI Weinshank R;  
PI WPI: 97-39326/36.  
DR P-PSDB; W24562.  
PT Nucleic acid molecule encoding galanin receptor GALT2 - agonists and  
PT antagonists of which are useful to treat eating disorders, pain and

PT Alzheimer's disease  
PS Claim 56; Fig 10; 236pp; English.  
CC This cDNA clone codes for a novel claimed galanin receptor subtype,  
CC designated GALT2 (see W24562). It was isolated from a human heart,  
CC cDNA library using primers based on rat GALT2 cDNA (see T79975).  
CC Also claimed are: (1) an isolated nucleic acid molecule (1) encoding  
CC a modified GALT2 having an amino acid deletion, replacement or  
CC addition in the third intracellular domain; (2) a purified GALT2  
CC protein; (3) a vector comprising (1); and (4) a cell comprising the  
CC vector of (3). A probe can be used to detect the expression of  
CC GALT2 by (claimed). GALT2 antagonists can be used to treat obesity,  
CC bulimia or Alzheimer's disease, while GALT2 agonists can be used to  
CC treat anorexia or pain, or to decrease nociception (claimed).  
CC Agonists and antagonists can also be used to treat other disorders,  
CC including cognitive disorders, sensory disorders, motion sickness,  
CC hypertension, convulsion/epilepsy, diabetes, glaucoma, reproductive  
CC disorders, gastric and intestinal ulcers, inflammation, immune  
CC presence of GALT2 on the surface of a cell, while transgenic  
CC mammals can be used to determine the physiological effects of  
CC varying levels of GALT2 activity (claimed).  
SQ Sequence 1365 BP: 184 A; 498 C; 424 G; 259 T;

Query Match 26.6%; Score 26.6; DB 1; Length 1365;  
Best Local Similarity 60.3%; Pred. No. 0.93;  
Matches 44; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 13 CTGATGAGCGCTTTGACGCCACATTCACACTACGCTGTGGAATCTCTGGGCATT 72  
DB 1278 CCGCTAAGTGGCTTTCAGCGCACATCAACCGTACGATGCTGCGCTTGGGCCCT 1219  
OY 73 GCCAGTACCTGCG 85  
DB 1218 GCCAGAGCGGCC 1206

RESULT 9  
V10329/c  
ID V10329 standard; cDNA; 1083 BP.  
AC V10329;  
DE 08-JUN-1998 (first entry)  
KW Human GALT2 receptor partial cDNA.  
KW Galanin receptor; GALT2; disease; treatment; screening; activity;  
KW antinociceptive activity; obesity; stroke; smooth muscle contraction;  
KW cardiovascular activity; diabetes; anorexia; pituitary hormone release;  
KW Alzheimer's disease; neuropathic pain; endocrine disorder; antibody;  
KW gene therapy; transgenic animal; human; ds.  
OS Homo sapiens.  
PN Key  
FT CDS Location/Qualifiers  
FT 1..997  
FT /\*tag= a  
FT /product= GALT2  
FT /note= "partial coding region of galanin recepto"  
FT MO9746681-A2.  
PD 11-DEC-1997.  
PF 05-JUN-1997; U09787.  
PR 03-JUN-1997; US-665034.  
PR 05-JUN-1996; US-665034.  
PA (FARB) BAYER CORP.  
PI Bloomquist BT, Cornfield LJ, Heeja Y, McCaleb ML;  
PI WPI: 98-042189/04.  
DR P-PSDB; W40137.  
PT Galanin receptor protein GALT2 and encoding polynucleotide - useful  
PT for e.g. identifying GALT2 receptor agonists or antagonists to  
PT regulate physiological responses e.g. to treat obesity and diabetes  
PS Claim 1; Fig 6; 44pp; English.  
CC This cDNA sequence encodes a novel human galanin receptor protein, GALT2.  
CC activity. Galanin has effects such as antinociceptive activity, smooth  
CC muscle contraction, cardiovascular activity, pituitary hormone release,  
CC cognition and increased food intake and identified drugs can regulate  
CC physiological responses associated with the GALT2 receptor, e.g.



CC GALR2 are useful in assays for GALR2 and to produce hybridisation  
CC probes to screen for similar receptors or for GALR2 in other species.  
SQ Sequence 3390 BP; 571 A; 1132 C; 1025 G; 660 T;

Query Match 26.6%; Score 26.6; DB 1; Length 3390;

Best Local Similarity 60.3%; Pred. No. 1.2; Mismatches 29; Indels 0; Gaps 0;

Matches 44; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 13 CTGATGAGCGCTTTGACGCCATTCACACTACGTTGGTGAATTCCTTCGGCATT 72

DB 2871 CCGCTAGAGTGGCTTTCAGGCCACATCAACCGTCAGAGATGCTGTCCCTTGCGGCCCT 2812

QY 73 GCCAGTACCTGCC 85

DB 2811 GCCAGAGACGGGCC 2799

RESULT 12

649/c V32649 standard; DNA; 3390 BP.

V32649; 24-NOV-1998 (first entry)

DE Human galanin receptor GALR2 genomic DNA.

KM Galanin receptor; GALR2 gene; human; ligand; obesity; anorexia;

OS Pain; cognitive disorder; therapy; G-protein coupled receptor; ss.

OS Homo sapiens. Location/Qualifiers

FT Key 305..2858

FT CDS /\*tag= a

FT exon /note= "contains an intron"

FT intron /tag= b

FT exon /number= 1

FT exon /tag= c

FT exon /number= 1

FT exon /tag= d

FT exon /number= 11

PN WO9829439-A1.

PD 08-JUL-1998.

PF 18-DEC-1997; U23890.

PR 27-DEC-1996; US-033851.

PA (MERI) MERCK & CO INC.

PI Sullivan K, Tan C;

DR WPI: 98-388037/33.

P-PSDB: W49003.

New galanin receptor, GALR2 - useful, e.g. to identify agonists and

antagonists, therapeutically to treat conditions involving excess or

insufficient galanin such as obesity

PS Example 6; Fig 7A-B; 57pp; English.

CC This is the DNA sequence of the human GALR2 gene that codes for a

novel galanin receptor, designated GALR2 (see W49003), that is a

member of the G-protein coupled receptor family. It was isolated

from an EMBL3 Sf6/77 human genomic library using as probe a partial

gene sequence that had been amplified from genomic DNA by PCR using

primers (see V32652-53) based on transmembrane regions of

CC somatostatin and somatostatin-related gene receptors. The human

CC GALR2 gene was mapped to chromosome 17q25. The invention provides

CC rat, human and mouse GALR2 sequences (see W49002-04). Also

CC provided are nucleic acids encoding these novel receptor sequences

CC and assays to identify ligands particular to GALR2. Such ligands

CC may be useful therapeutically e.g. to treat obesity or anorexia

CC disorders involving excess galanin or to treat pain or anorexia

CC involving insufficient galanin. Nucleic acids encoding GALR2 are

CC useful in assays for GALR2 and to produce hybridisation probes to

CC screen for similar receptors or for GALR2 in other species.

SQ Sequence 3390 BP; 571 A; 1132 C; 1025 G; 660 T;

Query Match 26.6%; Score 26.6; DB 1; Length 3390;

Best Local Similarity 60.3%; Pred. No. 1.2;

Matches 44; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 13 CTGATGAGCGCTTTGACGCCATTCACACTACGTTGGTGAATTCCTTCGGCATT 72

DB 2871 CCGCTAGAGTGGCTTTCAGGCCACATCAACCGTCAGAGATGCTGTCCCTTGCGGCCCT 2812

QY 73 GCCAGTACCTGCC 85

DB 2811 GCCAGAGACGGGCC 2799

RESULT 13

ID X13051 standard; DNA; 13574 BP.

AC X13051; 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome contig SEQ ID NO:114.

KM Enterococcus faecalis; contig; detection; Enterococcal infection;

OS vaccine; attenuation; computer readable medium; ds.

OS Enterococcus faecalis.

PN WO9850555-A2.

PD 12-NOV-1998.

PF 04-MAY-1998; U08985.

PR 14-NOV-1997; US-066009.

PR 06-MAY-1997; US-044031.

PR 16-MAY-1997; US-046555.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Dillon PJ, Kunsch CA;

DR WPI: 99-045171/04.

PT New isolated Enterococcus faecalis polynucleotides and polypeptides

PT - used to develop products for the detection of Enterococcus and for

PT use in vaccines for prevention or attenuation of Enterococcus

PT infection.

PS Claim 1; Page 706-713; 2084pp; English.

CC A computer readable medium has been developed which has recorded on it

CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.

CC X12938 to X13919 represent these nucleotide sequences which are primary

CC nucleotide sequences, also known as contigs. The computer-based system

CC can identify fragments of the Enterococcus faecalis genome with

CC commercial importance. The products can be used to detect the presence

CC of Enterococcus faecalis in samples. They can also be used for

CC diagnosing Enterococcal infection in an animal and monitoring

CC progression of disease, and for identifying agents which can be used to

CC modulate the growth or pathogenicity of Enterococcus faecalis, or

CC another related organism, in vivo or in vitro. In particular the

CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences

CC can be used in vaccines to prevent or attenuate an Enterococcal

CC infection.

SQ Sequence 13574 BP; 4658 A; 2376 C; 2922 G; 3608 T;

Query Match 25.2%; Score 25.2; DB 1; Length 13574;

Best Local Similarity 60.0%; Pred. No. 6.5;

Matches 42; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 4 TTTCTAGTGTGATGAGCGCTTTGACGCCATTCACACTACGTTGGTGAATTCCT 63

DB 3962 TTTCAGGTGCTTTGAGAGCGTGTGGGAAACACTTATATTAAGTGTGGACCAATAT 4021

QY 64 CTGGCATTCG 73

DB 4022 TTGGTATCG 4031

RESULT 14

X25011/c X25011 standard; DNA; 5002 BP.

AC X25011; 05-JUL-1999 (first entry)

DE Anthr- specific Esj2a gene and promoter.

KM Esj2a gene; promoter; pollen; anther dehiscence; male sterile;

OS transgenic plant; ss.

OS Arabidopsis thaliana.



CC of the nervous system. Neural stem cells are useful for the treatment of  
CC neurological disorders, especially for repair of accidentally induced  
CC trauma in the CNS or for correction of congenital or pathological  
CC diseases of the CNS. A patient with a neurological disorder can act as a  
CC self-donor. Cells may be isolated from the patient and either sorted to  
CC extract neuroblasts, or treated in order to differentiate neuroblasts,  
CC from specific or general precursors.  
SQ Sequence 2376 BP; 439 A; 772 C; 745 G; 420 T;

Query Match 24.4%; Score 24.4; DB 1; Length 2376;  
Best Local Similarity 54.4%; Pred. No. 7.4;  
Matches 49; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

OY 4 TTCTAGTGTGATGAGCGCTTTGAGCCACACTACACTACGCTGTGGTAATCTT 63  
| | | | | | | | | | | | | | | | | | | | | |  
Db 2100 TCTTTGGGTGGTGTGTTAATTATACCGCAATCCCTCTGAGCGGTGGAGTTATAT 2159  
| | | | | | | | | | | | | | | | | | | | | |  
OY 64 CTGGGCACTTGCAGTACCTGCCTCCAGGTA 93  
| | | | | | | | | | | | | | | | | | | | | |  
2160 CTGGGTTTGTAAATCTCTGTATCCGAGCA 2189

Search completed: September 18, 1999, 00:34:34  
Job time: 19034 sec

Gencore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:23 ; Search time 2825.05 Seconds  
(without alignments)  
69.823 Million cell updates/sec

Title: US-08-956-518a-99  
Perfect score: 100  
Sequence: 1 CTTCTCTGCTGCTGATGAG.....CTGCTCTCAGCTGATGCTGCA 100

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST:\*

1: em\_est1:\*

2: em\_est2:\*

3: em\_est3:\*

4: em\_est4:\*

5: em\_est5:\*

6: em\_est6:\*

7: em\_est7:\*

8: em\_est8:\*

9: em\_est9:\*

10: em\_est10:\*

11: em\_est11:\*

12: em\_est12:\*

13: em\_est13:\*

14: em\_est14:\*

15: em\_est15:\*

16: em\_est16:\*

17: em\_est17:\*

18: em\_est18:\*

19: em\_est19:\*

20: em\_est20:\*

21: em\_est21:\*

22: em\_est22:\*

23: em\_est23:\*

24: em\_est24:\*

25: em\_est25:\*

26: em\_est26:\*

27: em\_est27:\*

28: em\_est28:\*

29: em\_est29:\*

30: em\_est30:\*

31: em\_est31:\*

32: em\_est32:\*

33: em\_est33:\*

34: em\_est34:\*

35: em\_est35:\*

36: em\_est36:\*

37: em\_est37:\*

38: em\_est38:\*

39: em\_est39:\*

40: em\_est40:\*

41: em\_est41:\*

42: em\_est42:\*

43: em\_est43:\*

44: em\_est44:\*

45: em\_est45:\*

46: em\_est46:\*

47: em\_est47:\*

48: em\_est48:\*

49: em\_est49:\*

50: em\_est50:\*

51: em\_est51:\*

52: em\_est52:\*

53: em\_est53:\*

54: em\_est22:\*

55: em\_est23:\*

56: em\_est24:\*

57: em\_est25:\*

58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	77.2	77.2	409	25	W03952	W03952 za62c08.r1
2	56.4	56.4	560	26	W52861	W52861 zc03e02.r1
3	38.6	38.6	607	44	A1292581	A1292581 GH1518.5
4	27.6	27.6	569	47	A1520119	A1520119 LD40102.3
5	27.6	27.6	689	48	A1547741	A1547741 PN3.1.01
6	27.2	27.2	318	28	AA083907	AA083907 zm62h12.r
7	27.2	27.2	322	28	AA112074	AA112074 zm62h12.s
8	27.2	27.2	440	28	AA123934	AA123934 mp78e03.r
9	27.2	27.2	139	30	AA267622	AA267622 mz98d02.r
10	27.2	27.2	408	33	AA422190	AA422190 v44e03.r
11	27.2	27.2	178	37	AA726126	AA726126 v08e07.r
12	26.4	26.4	458	31	AA726126	AU010910 AU010910
13	25.8	25.8	374	22	R72681	R72681 y93d04.r1
14	25.8	25.8	216	29	AA138239	AA138239 mq78f03.r
15	25.8	25.8	522	46	A1450355	A1450355 mu28g03.x
16	25.6	25.6	350	26	W39186	W39186 zb35e08.r1
17	25.6	25.6	362	42	AU040432	AU040432 AU040432
18	25.6	25.6	274	49	AV016274	AV016274 AV016274
19	25.4	25.4	265	22	R35571	R35571 yh80g09.r1
20	25.4	25.4	456	22	R64451	R64451 y136g06.r1
21	25.4	25.4	484	49	A1635659	A1635659 t208d02.x
22	25.2	25.2	255	22	R28672	R28672 FI-119AD.22
23	25.2	25.2	475	25	N53430	N53430 y225g02.s1
24	25.2	25.2	444	26	W95996	W95996 ze08b09.r1
25	25.2	25.2	197	29	AA184843	AA184843 mu49c05.r
26	25.2	25.2	416	32	AA355266	AA355266 EST63893
27	25.2	25.2	272	32	AA360428	AA360428 EST69822
28	25.2	25.2	556	33	AA398679	AA398679 zt70e02.s
29	25.2	25.2	464	40	AA984510	AA984510 em88e06.s
30	25.2	25.2	407	41	A1004509	A1004509 O66h03.s
31	25.2	25.2	522	42	A1147987	A1147987 g963b07.s
32	25.2	25.2	406	44	A1312703	A1312703 qp83d02.x
33	25.2	25.2	188	49	A1627046	A1627046 mu49c05.y
34	25.0	25.0	500	35	AA590268	AA590268 vml19e02.r
35	25.0	25.0	614	43	A1177061	A1177061 EST720668
36	25.0	25.0	795	44	A1246505	A1246505 qm64a10.x
37	25.0	25.0	427	46	A1437303	A1437303 fd40a12.y
38	25.0	25.0	472	47	A1490165	A1490165 EST248796
39	25.0	25.0	471	47	A1490177	A1490177 EST248808
40	24.8	24.8	277	32	AA377412	AA377412 EST90023
41	24.8	24.8	277	42	A115208	A115208 UT-R-BTO-
42	24.8	24.8	623	47	A1487007	A1487007 EST245329
43	24.6	24.6	340	21	F11432	F11432 HSC2WH081.n
44	24.6	24.6	303	22	R46208	R46208 y153e02.s1
45	24.6	24.6	411	24	N34208	N34208 yx77g03.r1

## ALIGNMENTS

RESULT 1

LOCUS W03952 409 bp mRNA EST 19-APR-1996

DEFINITION za62c08.r1 Soares fetal liver spleen INF18 Homo sapiens cDNA clone IMAGE:297134.5' similar to gb:X70297 NEURONAL ACETYLCHOLINE RECEPTOR PROTEIN, ALPHA-7 CHAIN (HUMAN);, mRNA sequence.

ACCESSION W03952

NID g1273817

REFERENCE	TITLE	COMMENTS
<p> <b>REFERENCE</b>  <b>AUTHORS</b>            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;            Eutheria; Primates; Catarrhini; Hominoidea; Homo.            1 (bases 1 to 560)            Haller, L., Clark, N., Dubouche, T., Elliston, K., Hawkins, M.,            Holman, M., Hillman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,            Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,            Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and            Wilson, R.            The Mashu-Merk EST Project            Unpublished (1995)            On Jan 24, 1995 this sequence version replaced gi:534363.         </p>	<p> <b>TITLE</b>  <b>JOURNAL</b>  <b>COMMENT</b>            Contact: Wilson RK            Washington University School of Medicine            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108            Tel: 314 286 1800            Fax: 314 286 1810            Email: est@wustl.wustl.edu            This clone is available royalty-free through LNL; contact the            IMAGE Consortium (info@image.lnl.gov) for further information.            Insert Length: 1753 Std Error: 0.00            Seq primer: mob.RS8A+RT            High quality sequence stop: 408.            Location/Qualifiers            1. 560            /organism="Homo sapiens"            /db_xref="GDB:1258900"            /db_xref="taxon:9606"            /clone="IMAGE:21242"            /clone_lib="Soares_parityroid_tumor_NbHPA"            /tissue_type="parathyroid tumor"            /dev_stage="adult"            /lab_host="DH10B (ampicillin resistant)"            /note="Organ: parathyroid gland; Vector: pTT73D            (Pharmacia) with a modified polylinker; Site_1: Not I;            Site_2: Eco RI; 1st strand cDNA was primed with a Not I -            oligo(dT) primer            15'-         </p>	<p> <b>FEATURES</b>  <b>Source</b>            BASE COUNT 128 a 137 c 137 g 152 t 6 others            ORIGIN            Query Match 56.4%; Score 56.4; DB 26; Length 560;            Best Local Similarity 95.2%; Pred. No. 3.3e-10;            Matches 79; Conservative 0; Mismatches 2; Indels 2; Gaps 2;            Oy 8 TAGGCTGATGAGGCGCTTGACGCCACATTCACACTAGCTGTTGGATCTTCTGG 67            Db 84 TATTCCTATATAGGCGCTTGAC-NCAATTCACACACTAGCTGTTGGATCTTCTGG 142            Oy 68 GCATTCACACTAGCTGCTCCAG 90            Db 143 GCATTCG-CAGTACTGCTCCAG 164         </p>



KEYWORDS EST.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;  
Muscomorpha; Ephydriidae; Drosophilidae; Drosophila.  
1 (bases 1 to 607)  
REFERENCE Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,  
Brockstein, P., Lewis, S. and Rubin, G. M.  
TITLE BDGP/HMI Drosophila Est Project  
JOURNAL Unpublished (1997)  
COMMENT On Jun 18, 1996 this sequence version replaced gi:1366531.

Contact: Harvey, D.  
G. M. Rubin-Molecular and Cell Biology  
University of California Berkeley  
539 LSA, Berkeley, CA 94720-3200, USA  
Fax: 510 643 9947  
Email: [http://www.fruitfly.org/EST\\_estrefruitfly.berkeley.edu](http://www.fruitfly.org/EST_estrefruitfly.berkeley.edu)  
hit genomic sequence AC005890  
plate: 155 row: B column: 6  
high quality sequence: stop: 465.

URES	Location/Qualifiers
source	1. .607

BASE COUNT	157 a	151 c	162 g	137 t
ORIGIN				

RESULT	4			
LOCUS	AI520119			
DEFINITION	AI520119	569 bp	mRNA	EST
ACCESSION	LD040102.3prime	LD	Drosophila melanogaster	embryo POT2 Drosophila
VERSION	AI520119			
KEYWORDS	94425973			
SOURCE	AI520119.1	GI:4425973		
ORGANISM	EST.			
REFERENCE	fruit fly.			
AUTHORS	Drosophila melanogaster			
TITLE	Euarystia, Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
JOURNAL	Pterygota, Neoptera; Endopterygota; Diptera; Brachycera;			
COMMENT	Muscomorpha, Ephydroidea; Drosophilidae; Drosophila..			
	1 (baes 1 to 569)			
	Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,			
	Brokstein, P., Lewis, S. and Rubin, G.M.			
	BDGP/HMT Drosophila EST Project			
	Unpublished (1997)			
	On Jun 5, 1998 this sequence version replaced gi:3188811.			
	Contact: Harvey, D.			

G.M. Rubin-Molecular and Cell Biology  
University of California Berkeley  
539 USA, Berkeley, CA 94720-3200, USA  
Fax: 510 643 9947  
Email: [hitp@www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](mailto:hitp@www.fruitfly.org/EST_est@fruitfly.berkeley.edu)  
Based upon the presence of a XhoI site followed by a run of 14 or more T residues at the beginning of the sequence, this clone probably contains an inverted insert. The resulting Poly-T sequences has been removed. hit genomic sequence AC004728  
Plate: 401 row: A column: 2  
High quality sequence: stop: 494.  
Location/Qualifiers  
1..569 - [fructose 1,6-bisphosphate](#)

BASE COUNT	111 a	107 c	170 g	180 t	1 others
ORIGIN					

Query Match	27.68;	Score 27.6;	DB 47;	Length 569;
Best Local Similarity	55.18;	Pred. No. 5.6;		
Matches 54;	Conservative 0;	Mismatches 44;	Indels 0;	Gaps 0.

QY 3 GATTGTATGCGAATAGACGGTTTGGACCCCATTCACACATTAAGCGTGGTAATCT 62  
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
 Db 400 GTTATGTGCTGCTGATGCTGGCTCCGGCCATCCACACACGGCCCCCGCATTTCCTTGGGG 458  
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
 QY 63 TCTGGGCAATTGCCAGTAACTTACCTGCTTCCAGATTAAGTGA 100  
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
 Db 460 TGGCGGCGATGGCGGCTGCTCTTCTATACGCGCTGTGCA 497  
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

**Contact:** Guyang Matthew Huang  
 Leroy Hood  
 University of Washington  
 Department of Molecular Biotechnology, Box 357730, University of  
 Washington, Seattle, WA 98195  
 Tel: 5106280100  
 Fax: 5106280108  
 Email: [huanggm@yahoo.com](mailto:huanggm@yahoo.com).  
 Location/Qualifiers  
     1. 689  
     /organism="Homo sapiens"  
     /db\_xref="taxon:9606"  
     /map="974810; 19; Chromosome 11; 21q"  
     /clone.lib="mynorm"

	Query Match	27.4%;	Score 27.4;	DB 48;	Length 689;
	Best Local Similarity	59.7%;	Pred. No. 7.1;		
	Matches	46;	Conservative	0;	Mismatches 31;
					Indels 0;
					Gaps 0;
Qy	24	TTTGAGCCACATTCACACTACACTGTGGGAATCTTCTGGGCATTCGACGATCTCG	83		
Db	26	TTTGCCACAGACGGAAACAACCACTGCTTGGGATTCGTCTCAACATGATAGAGACAG	85		
	84	CCCTCCAGGTAGCTGCA	100		
	86	CCGACATCGAAGGATCA	102		

## FEATURES

### source

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estewartson.wustl.edu  
 WARNING: There is evidence that suggests that the 384-well parent plate of this clone contains both human and mouse derived clones. Thus, the origin of this clone is uncertain. This caution should be kept in mind should you use this clone.

This clone is available royalty-free through LINT ; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 Seq primer: -28M1 rev2 from Amersham.

Location/Qualifiers  
 1..318  
 /organism="Homo sapiens"  
 /db\_xref="GDB:391979"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:530279"  
 /clone\_lib="Stratagene fibroblast (#937212)"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /note="vector: pBluescript SK-; site\_1: EcoRI, site\_2: XhoI, cloned unidirectionally. Primer: oligo dT, W38 cell line. Average insert size: 0.8 kb. Unt-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGACGACGAG 3' -3' adaptor

	Query Match	Best Local Similarity	27.28;	Score 27.2;	DB 28;	Length 318.
	Matches	41; Conservative	64.18;	Pred No. 6.3;	Mismatches 23;	Indels 0; Gaps 0;
QY	24	TTTTACCCACATTCACACATCAAGTGTGTGATTCCTGTGGGCATTGCCAGTACCTG	83			
Db	179	TTTCTGTGAATTAACATCACTACGAGTAGAGGGCTATCCACTGCCCATTTGTCACGTCTC	120			
QY	84	CCTC	87			
Db	119	CCTC	116			

RESULT	7
AAL12074	
LOCUS	
DEFINITION	AA112074 322 bp MRNA EST 23-DEC-1997
ACCESSION	z62n12.s1 Striatagene fibroblast (#937212) Homo sapiens cdna clone
NID	IMAGE:530279 3', mRNA sequence.
VERSION	AA112074
KEYWORDS	g1664253
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
ATTNORS	Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 322)
TITLE	Miller,L., Lannon,G., Becker,M., Donald,M.F., Chiappelli,B., Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Marlis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaslis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
JOURNAL	Generation and analysis of 280,000 human expressed sequence tags Genome Res. 6 (9), 807-828 (1996) MEDLINE 97044478
COMMENT	On May 9, 1995 this sequence version replaced g1:802978.

**FEATURES**  
**source**

```

FEATURES
Source
This clone is available royalty-free through LINT; contact the
IMAGE Consortium (info@image.ln1.gov) for further information.
Seq primer: -40M13 fwd from Amer sham.
Location/Qualifiers
1..322
/organism="Homo sapiens"
/db_xref="GDB:3919790"
/db_xref="taxon:9606"
/map="1"
/clone="IMAGE:530279"
/clone_lid="Stratagene fibroblast (#937212)"
/lab_host="SOLR cells (Kanamycin resistant)"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dT. W138
-5' adaptor sequence: 0.8 kb; Uni-ZAP XR Vector;
sequence: 5' CTGCAAGTGTGTTTTTTTTTTTTTTT 3'"
77 a 75 c 68 g 102 t
BASE COUNT

```



BASE COUNT 43 a 30 c 35 g 31 t  
ORIGIN

Query Match  
Best Local Similarity 27.2%; Score 27.2; DB 30; Length 139;  
Matches 41; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 24 TTTCAGCCACATTCACACTAGCTGTGGATTTCTTCTGGGCAATGGCAGTACCTG 83  
DB 85 TTTCGTGCAACATTAACACTCAGAGTGGATTCACACTGCTCCTACTGCTC 26  
QY 84 CCTC 87  
DB 25 CTTC 22

LOCUS 22190/c 408 bp mRNA EST 16-OCT-1997  
DEFINITION v144603.r1 Soares mouse NBMH Mus musculus cDNA clone IMAGE:846652  
ACCESSION AA422190  
NID 92101108  
VERSION AA422190.1 GI:2101108  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 408)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Giesel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE The WashU-HMNI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Nov 29, 1993 this sequence version replaced gi:637644.

CONTACT: Marra M/Mouse EST Project  
WashU-HMNI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:498804  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 356.  
Location/Qualifiers  
1. 408  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_image="IMAGE:846652"  
/clone\_lib="Soares mouse NBMH"  
/sex="male"  
/tissue\_type="heart"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTCACATCTGAGTGGAGCGGCGGAGAGTTTCTTTTCTTTTCTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not  
I and Eco RI sites of the modified pT73 vector. RNA  
provided by Dr. Minoru Ko, Wayne State Univ. Library  
constructed and normalized by Bento Soares and M. Fatima  
Bonaldo."

BASE COUNT 99 a 84 c 109 g 116 t  
ORIGIN

Query Match  
Best Local Similarity 27.2%; Score 27.2; DB 33; Length 408;  
Matches 41; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 24 TTTCAGCCACATTCACACTAGCTGTGGATTTCTTCTGGGCAATGGCAGTACCTG 83  
DB 339 TTTCGTGCAACATTAACACTCAGAGTGGATTCACACTGCTCCTACTGCTC 280  
QY 84 CCTC 87  
DB 279 CTTC 276

RESULT 11  
LOCUS AA726126 178 bp mRNA EST 02-JAN-1998  
DEFINITION vU88C07.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone  
IMAGE:1209228 5', mRNA sequence.  
ACCESSION AA726126  
NID 92743833  
VERSION AA726126.1 GI:2743833  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 178)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Giesel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE The WashU-HMNI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On May 18, 1995 this sequence version replaced gi:811372.

CONTACT: Marra M/Mouse EST Project  
WashU-HMNI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:645572  
Putative full length read  
vector to vector length is 536  
Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 162.  
Location/Qualifiers  
1. 178  
/organism="Mus musculus"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/map="7"  
/clone\_image="1209228"  
/clone\_lib="Stratagene mouse skin (#937313)"  
/sex="females"  
/tissue\_type="whole skin"  
/dev\_stage="11 weeks old"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: skin; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Whole skin from 11 week old C57BL/6 female mice.  
Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'  
adaptor sequence: 5' CTCGACTTTTCTTTTCTTTTCTTTT 3'."  
sequence: 5' CTCGACTTTTCTTTTCTTTTCTTTT 3'."

BASE COUNT

50 a 38 c 46 g 44 t



RESULT 14  
 AA138239 216 bp mRNA EST 09-FEB-1997  
 LOCUS  
 DEFINITION mq78f03.f1 Stragatene mouse melanoma (#937312) Mus musculus cDNA  
 clone IMAGE:584861 5', mRNA sequence.  
 ACCESSION AA138239  
 MID 91700442  
 VERSION AA138239.1 GI:1700442  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 216)  
 Marti, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Scheinberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.  
 TITLE The WashU-HM1 Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1394413.

CONTACT: Maria M/Mouse EST Project  
 WashU-HM1 Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 Seq primer: -28ml3 rev1 ET from Amersham  
 High quality sequence stop: 66.  
 Location/Qualifiers  
 1..216  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_image="584861"  
 /clone\_lib="Stratagene mouse melanoma (#937312)"  
 /tissue\_type="melanoma"  
 /dev\_stage="M2 cells"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: skin; Vector: pBluescript SK-; Site: 1;  
 EcoRI; Site: 2; XhoI; Cloned unidirectionally. Primer:  
 Oligo dT. From M2 cells, a highly metastatic derivative of  
 the K-1735 (mouse) melanoma. Average insert size: 1.0 kb;  
 Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCGACGAG  
 3' -3' adaptor sequence: 5' CTCGAGCTTTTCTTTTCTTTT 3' "

BASE COUNT 83 a 39 c 39 g 55 t  
 ORIGIN

Query Match 25.8%; Score 25.8; DB 29; Length 216;  
 Best Local Similarity 58.4%; Pred. No. 17;  
 Matches 45; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

9 AGTGTGATGAGCGCTTGGACGACATTCACATACGTTGGTGAATTCCTTGGG 68  
 Db 115 AATGCTGATGTGTGTTATGCGACATCTCCAACTACAGACTCAGTACATATTCAA 174  
 Oy 69 CATTGCCAGTACTGCC 85  
 Db 175 CAAGAAGAGTGTCAATC 191

RESULT 15  
 AA150355 522 bp mRNA EST 09-MAR-1999  
 LOCUS  
 DEFINITION mu28g03.x1 Soares 2NbMT Mus musculus cDNA clone IMAGE:640756 3',

mRNA sequence.  
 ACCESSION AA150355  
 MID 94296423  
 VERSION AA150355.1 GI:4296423  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 522)  
 Marti, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
 Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,  
 Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
 Waterston, R. and Wilson, R.  
 TITLE The WashU-NCI Mouse EST Project 1999  
 JOURNAL Unpublished (1999)  
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3189570.

CONTACT: Maria M/WashU-NCI Mouse EST Project 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 This clone was previously sequenced on the 5' end only, this new  
 data is from the 3' end  
 High quality sequence stop: 449.  
 Location/Qualifiers  
 1..522  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_image="640756"  
 /clone\_lib="Soares 2NbMT"  
 /sex="male"  
 /tissue\_type="Thymus"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site: 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5',  
 TGTACCAATCTAATGAGGCGCGCGCTTTTCTTTTCTTTTCTTTTCTTTT  
 3']; double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. RNA  
 provided by Dr. Bertrand Jordan. Library went through two  
 rounds of normalization, and was constructed by Bento  
 Soares and M. Fatima Bonaldo."

BASE COUNT 174 a 106 c 103 g 139 t  
 ORIGIN

Query Match 25.8%; Score 25.8; DB 46; Length 522;  
 Best Local Similarity 58.4%; Pred. No. 24;  
 Matches 45; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

9 AGTGTGATGAGCGCTTGGACGACATTCACATACGTTGGTGAATTCCTTGGG 68  
 Db 123 AATGCTGATGTGTGTTATGCGACATCTCCAACTACAGACTCAGTACATATTCAA 182  
 Oy 69 CATTGCCAGTACTGCC 85  
 Db 183 CAAGAAGAGTGTCAATC 199

Search completed: September 17, 1999, 21:28:26  
 Job time: 14290 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:30 ; Search time 2825.05 Seconds

(without alignments)  
481.080 Million cell updates/sec

Title: US-08-956-518a-102

Sequence: 1 ACCCCCTTCCAGCGCGTAG.....CAAGATTCTCTACATCG 689

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database: EST.\*

1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: em\_est20: \*  
21: em\_est21: \*  
22: em\_est22: \*  
23: em\_est23: \*  
24: em\_est24: \*  
25: em\_est25: \*  
26: em\_est26: \*  
27: em\_est27: \*  
28: em\_est28: \*  
29: em\_est29: \*  
30: em\_est30: \*  
31: em\_est31: \*  
32: em\_est32: \*  
33: em\_est33: \*  
34: em\_est34: \*  
35: em\_est35: \*  
36: em\_est36: \*  
37: em\_est37: \*  
38: em\_est38: \*  
39: em\_est39: \*  
40: em\_est40: \*  
41: em\_est41: \*  
42: em\_est42: \*  
43: em\_est43: \*  
44: em\_est44: \*  
45: em\_est45: \*  
46: em\_est46: \*  
47: em\_est47: \*  
48: em\_est48: \*  
49: em\_est49: \*  
50: em\_est50: \*  
51: em\_est51: \*  
52: em\_est52: \*  
53: em\_est53: \*

54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	128.4	18.6	409	25	W03952
2	115.8	18.8	560	26	W52861
3	106	15.4	170	25	N83152
4	92.2	13.4	510	54	HS0011948
5	68.6	10.0	335	32	AA333244
6	49.4	7.2	607	44	A1292581
7	46.8	6.8	487	39	AA861176
8	42.2	6.1	274	34	AA460018
9	42.2	6.1	436	35	AA844642
10	42.2	6.1	396	45	A1367541
11	42.2	6.1	501	49	A1655990
12	41.8	6.1	521	53	HS0003697
13	40.8	5.9	494	37	AA697326
14	39.4	5.7	439	47	A1492967
15	39	5.7	557	43	A1192683
16	38.6	5.6	369	29	AA156428
17	37.8	5.5	354	26	W86837
18	37.8	5.5	790	48	A1555069
19	37	5.4	473	33	AA407186
20	37	5.4	446	44	A1317059
21	36.6	5.3	319	33	AA408177
22	36.4	5.3	400	50	A1681298
23	36.4	5.3	399	51	AU067986
24	36	5.2	435	42	A1085561
25	36	5.2	654	45	A1346522
26	36	5.2	373	47	A1462417
27	36	5.2	373	47	A1462417
28	36	5.2	415	24	H81299
29	35.6	5.2	289	37	AA702650
30	35.6	5.2	391	48	A1586795
31	35.6	5.2	605	40	AA979965
32	35.4	5.1	317	46	A1443300
33	35.4	5.1	638	26	W27222
34	35.2	5.1	232	20	D22064
35	35	5.1	232	20	D22064
36	35	5.1	216	20	D22108
37	35	5.1	216	20	D24865
38	35	5.1	443	27	AA040321
39	35	5.1	482	39	AA894793
40	35	5.1	415	45	A1363984
41	35	5.1	381	46	AA824494
42	35	5.1	482	49	A1660583
43	34.8	5.1	528	49	A1621763
44	34.8	5.1	630	49	A1622084
45	34.8	5.1	571	49	A1622684

## ALIGNMENTS

RESULT 1  
LOCUS W03952  
DEFINITION W03952 409 bp. mRNA  
IMAGE: 297134 5' similar to gb:X70297 NEURONAL ACETYLCHOLINE RECEPTOR PROTEIN, ALPHA-7 CHAIN (HUMAN);, mRNA sequence.  
ACCESSION W03952  
NID 91275817

EST 19-APR-1996

VERSION	W03952.1	GI:1275817
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
AUTHORS	Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
	1 (bases 1 to 409)	
	Hillier,L., Clark,N., Dubouche,T., Eiliston,K., Hawkins,M.,	
	Holman,M., Hulman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,	
	Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,	
	Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and	
	Wilson,R.	
TITLE	The WashU-Merck EST Project	
JOURNAL	Unpublished (1995)	
COMMENT	Contact: Wilson RK Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu This clone is available royalty-free through LNC; contact the IMG Consortium (info@image.llnl.gov) for further information. Seq primer: mob.REGA+Er High quality sequence stop: 263.	
FEATURES	Location/Qualifiers	
Source	1..409	
	/organism="Homo sapiens"	
	/db_xref="GDB:1242035"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:297134"	
	/clone_lib="Soares fetal liver spleen INFLS"	
	/sex="male"	
	/dev_stage="20 week-post conception fetus"	
	/lab_host="DH10B (ampicillin resistant)"	
	/note="Organ: Liver and Spleen; Vector: pTV73D (Pharmacia)	
	with a modified polylinker; Site.1: Pac I; Site.2: Eco RI;	
	1st strand cDNA was primed with a Pac I - oligo(dT) primer	
	{5' AACTGGAGAAATTAATTAAAGCTTTTTTTTTTTTTTTTTT 3'},	
	double-stranded cDNA was ligated to Eco RI adaptors	
	(pharmacia), digested with Pac I and cloned into the Pac I	
	and Eco RI sites of the modified pTV73 vector. Library	
	went through one round of normalization. Library	
	constructed by Bento Soares and M.Fatima Bonaldo."	
BASE COUNT	95 a 80 c 75 g 145 t	14 others
OP'TIN		
Query Match	18.6%; Score 128.4; DB 25; Length 409;	
Best Local Similarity	92.98; Pred. No. 1e-21;	
Matches 143; Conservative	0; Mismatches 10; Indels 1; Gaps 1	
OY	534 AATTCGAATTCGTATCA-GCATTGTGATGACCTGCACAACCTGCATATTCGTATGAG 592	
Db	157 AGTCCATTCTCTATCACGCCCATTTMTNGATACCTGCACAACCTGCATATTCGTATGAG 216	
OY	593 CGCTTGACGCCACATTCACACTAACGCTGTGTGTAATCTTCTGGCGATTCGCCAGTAC 652	
Db	217 CGCTTGACNCCACATTCACACTAACGCTGTGTGTAATCTTCTGGCGATTCGCCAGTAC 276	
OY	653 CTGCTCCAGCGCATATTCAGAAGTTCCTCGCTACA 686	
Db	277 CTGCTCCAGGNATTAATTCAGAAGTTCCTCGCTACA 310	
RESULT	2	
LOCUS	W52861	
DEFINITION	M52861 560 bp mRNA EST 10-OCT-1996	
ACCESSION	W52861	
NID	91350351	

	VERSION	M52861.1	GI:1350351
	KEYWORDS	EST.	
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
	Euarchyonta; Primates; Catarrhini; Hominoidea; Homo.		
REFERENCE	1 (bases 1 to 560)		
AUTHORS	Hillier L., Clark N., Dubucque T., Elliston K., Hawkins M., Holman M., Hultman M., Kucabara T., Le M., Lennon G., Marra M., Parsons J., Rinkin L., Rohlfing T., Soares M., Tan F., Trevaskis E., Waterston R., Williamson A., Woldmann P. and Wilson R.		
TITLE	The WashU-Merck EST Project		
JOURNAL	Unpublished (1995)		
COMMENT	On Jan 24, 1995 this sequence version replaced gi:634363.		
	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@waton.wustl.edu  This clone is available royalty-free through LNL ; contact the IMAGE Consortium ( <a href="#">info@image.lnl.gov</a> ) for further information. Insert Length: 1753 Std Error: 0.00 Seq primer: mob.REGA+ET High quality sequence stop: 408. Location/Qualifiers 1..560 /organism="Homo sapiens" /db_xref="GDB:11258900" /db_xref="taxon:9606" /clone="IMAGE:321242" /cloae_lib="Soares parathyroid_tumor_NbhPA" /tissue_type="parathyroid tumor" /dev_stage="adult" /lab_host="DH10B (ampicillin resistant)" /note="Organ: Parathyroid gland; Vector: pRT73D (Pharmacia) with a modified polylinker; Site_1: Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'- TGTTCCCAATCTGAAGTCGGACGCCGCACCACAATTGTCTTTTTTTTTTTTTTTTTTTT -T-3'] , double-stranded CDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRT733 vector (Pharmacia). Library went through one round of normalization to a Cot = 5 library constructed by Benton, Soares and M.Felina Bonaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH."		
BASE COUNT	128 a	137 c	137 g    152 t    6 others
ORIGIN	NH.""		
	Query Match	16.8%;	Score 115.8; DB 26; Length 560;
	Best Local Similarity	95.5%;	Pred. No.m.1.3e-18;
	Matches 150; Conservative	0;	No.matches 4; Indels 3; Gaps 3
OY	533 CAATTCGAATTGCTAAATCGCAGCATTTCTGGATAGCGCAAACATGCGATAATGCTGATGAG	592	
Dd	37 CAGTTCGAATGCAATGCAATCGCACGACTTTGTAATGCTGCAACATGCTGATGATGAG	96	
OY	593 CGGTTTAGCGGCACATTCACACACTAACGCTGTGGTAATTTCTTGGGCAATTCACGATAC	652	
Dd	97 CGCTTTAGC-NCAATTCACACACTAACGCTGTGGTAATTTCTTGGGCAATTC-CAGTAC	154	
OY	653 CTGCCTCAGGCATATTCAGAAGTTCCTGCTACTACG	689	
Dd	155 CTCGCTCCA-GNATATTCAGAAGTTCCTGCTACTACG	190	



RESULT 3  
 N83152 170 bp mRNA EST 02-APR-1996  
 LOCUS N83152  
 DEFINITION K4329F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA  
 clone K4329 5' similar to R KAPPA B(NON-EXACT), mRNA sequence.  
 ACCESSION N83152  
 NID 91441866  
 VERSION N83152.1 GI:1441866  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 170)  
 AUTHORS Liew, C.C.  
 TITLE CDNA's from fetal heart (1996)  
 JOURNAL Unpublished (1996)  
 COMMENT On Apr 14, 1993 this sequence version replaced g1:502850.

CONTACT: Liew CC  
 Department of Laboratory Medicine and Pathobiology  
 University of Toronto  
 Banting Institute, 100 College St., Toronto, Ontario, M5G1L5  
 Tel: 4169788758  
 Fax: 4169785650  
 Email: liewc@utcc.utoronto.ca  
 Seq primer: GAATTAACTTCACTCAAGG.  
 FEATURES  
 source  
 1..170  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="K4329"  
 /clone\_1lb="Human fetal heart, Lambda ZAP Express"  
 /lab\_host="E. coli XL1-Blue"  
 /note="Vector: Lambda ZAP Express; Site\_1: EcoRI; Site\_2: XhoI; mRNA was purified from human fetal hearts (8-10 weeks). cDNA was synthesized using a XhoI-0.150 dt adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into predigested lambda ZAP Express."

BASE COUNT 43 a 47 c 46 g 34 t  
 ORIGIN

Query Match 15.4%; Score 106; DB 25; Length 170;  
 Best Local Similarity 82.9%; Pred. No. 2.1e-16;  
 Matches 121; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

C 268 GGCTCTTTCCGCGCCCTCCGCGGAGGTGAGGGAAGATGTCATGTCAGGGTTCA 327  
 L 9 GGCTCCCTTCACAGCCCTCCGCGGAGATGAGGGAAGATGTCATGTCAGGATTC 68  
 QY 328 AGGCCAAACGAGTACTGCTCTATCTTCCAGAGAACACAGAGCCACAGCCCGCGC 387  
 DB 69 AGGCCAAACGAGTACTGCTCTATCTTCCAGAGAACACAGAGCCACAGCCCGCGC 387  
 QY 388 TCAGGCCCGACCGACCATTAAGATT 413  
 DB 129 TCAGGTCACCTTCAACATGAGATT 154

RESULT 4  
 HSM011948 standard; RNA; EST; 510 BP.  
 ID HSM011948  
 AC AL047098;  
 SV AL047098.1  
 NI e1405574  
 XX  
 DT 12-MAR-1999 (Rel. 59, Created)  
 DT 12-MAR-1999 (Rel. 59, last updated, Version 1)

XX Homo sapiens mRNA; EST DKEZP586B2218\_r1 (from clone  
 DE DKEZP586B2218)  
 XX EST; expressed sequence tag.  
 KW  
 XX  
 OS Homo sapiens (human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;  
 CC Primates; Catarrhini; Homidae; Homo.  
 XX  
 [1]  
 RA Koehler K., Beyer A., Mewes W., Gassenhuber J., Wiemann S.;  
 RP 1-510  
 RT Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RL MIPS, Am Klopferst 18a D-82152 Martinsried, GERMANY  
 CC Clone from S. Wiemann, sequenced by BMF within the CDNA  
 CC sequencing consortium of the German Genome Project  
 CC No SI sequence available  
 CC This clone is available at the RZPD in Berlin  
 CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
 CC Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de  
 FH  
 FH Key  
 Location/Qualifiers  
 1..510  
 /db\_xref="taxon:9606"  
 /organism="Homo sapiens"  
 /clone="DKEZP586B2218"  
 /clone\_1lb="586 (synonym: hute1). Vector pSport1; host  
 /dev-stage="adult"  
 /tissue-type="uterus"  
 XX  
 SO Sequence 510 BP; 118 A; 154 C; 150 G; 88 T; 0 other;

Query Match 13.4%; Score 92.2; DB 54; Length 510;  
 Best Local Similarity 73.7%; Pred. No. 6.5e-13;  
 Matches 154; Conservative 0; Mismatches 33; Indels 22; Gaps 2;

QY 191 GCGGGGAGAGTGGAGCCGCGAGAGCTCGGCC---GGGGCCCGCTGTGCGCCGCGC 247  
 DB 3 GCTGACGAGAGCGGACCGCGAGAGCGCGCCAGCGCCCGCGGGCGGTGCGCGC 62  
 QY 248 CAGGACAGCGGCTCGGAGCT-----GGTCTCTTTCCGCGCCCTTC 288  
 DB 63 CGTGAAGGCGGCTCGGCGCGCGGCTCCCTCCGCGCCCGCTCCCTCCGCGCCCTC 122  
 QY 289 CCGCGGAGAGTGGAGGGAAGATGTCATGTCAGGGTTCAAGGCAACGAGTACTGG 348  
 DB 123 CCGCGGAGATGAGGGAAGATGTCGCTCAGGCTCAAGGCCAGCTGAAGTCTCTGG 182  
 QY 349 CCTATGCTTCAGAGAACACAGAGCA 377  
 DB 183 CGTCACTCTCGACAGAACACAGAGCA 211

RESULT 5  
 AA333244  
 LOCUS AA333244  
 DEFINITION AA333244 335 bp mRNA EST 21-APR-1997  
 EST37317 Embryo, 8 week I Homo sapiens cDNA 5' end similar to  
 similar to Interleukin-2 receptor, alpha chain, kappa B binding  
 protein, mRNA sequence.  
 ACCESSION AA333244  
 NID 91985508  
 VERSION AA333244.1 GI:1985508  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homidae; Homo.





/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from Clontech  
Laboratories, Inc., and primed with a Not I - oligo(dT)  
primer [5].  
TTCACATGCTAATCCAGCATTTGTGATAGCTGCAAACTCGGATAT 582  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization to cots, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 115 a 96 c 103 g 122 t  
Query Match 6.1%; Score 42.2; DB 39; Length 436;  
Best Local Similarity 93.6%; Pred. No. 0.84; 3; Indels 0; Gaps 0;  
Matches 44; Conservative 0; Mismatches 0;

QY 536 TTCACATGCTAATCCAGCATTTGTGATAGCTGCAAACTCGGATAT 582  
Db 266 TTCACATGCTAATCCAGCATTTGTGATAGCTGCAAACTCGGATAT 220

RESULT 10  
A1367541/c 396 bp mRNA EST 15-FEB-1999  
LOCUS qv39c12.x1 NCI\_CGAP\_Ut2 Homo sapiens cDNA clone IMAGE:1989142 3'  
DEFINITION mRNA sequence.  
ACCESSION A1367541  
NID 94137286  
VERSION A1367541.1 GI:4137286  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Mammalia;  
Eukaryota; Metazoa; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 396)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

FEATURES  
source  
1. 396  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1989142"  
/clone\_id="NCI\_CGAP\_Ut2"  
/tissue\_type="moderately-differentiated endometrial  
adenocarcinoma, 3 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: PCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.85 Kb. Life Technologies catalog #:  
11539-012"

BASE COUNT 98 a 86 c 97 g 115 t

## ORIGIN

Query Match 6.1%; Score 42.2; DB 45; Length 396;  
Best Local Similarity 93.6%; Pred. No. 0.81; 3; Indels 0; Gaps 0;  
Matches 44; Conservative 0; Mismatches 0;

QY 536 TTCACATGCTAATCCAGCATTTGTGATAGCTGCAAACTCGGATAT 582  
Db 277 TTCACATGCTAATCCAGCATTTGTGATAGCTGCAAACTCGGATAT 231

RESULT 11  
A1655990/c 501 bp mRNA EST 04-MAY-1999  
LOCUS t12c03.x1 NCI\_CGAP\_GC6 Homo sapiens cDNA clone IMAGE:2243428 3'  
DEFINITION mRNA sequence.  
ACCESSION A1655990  
NID 94739969  
VERSION A1655990.1 GI:4739969  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Mammalia;  
Eukaryota; Metazoa; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 501)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
JOURNAL Unpublished (1997)  
COMMENT On Jun 5, 1998 this sequence version replaced gi:3188057.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco  
High quality sequence stop: 458.  
Location/Qualifiers

FEATURES  
source  
1. 501  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2243428"  
/clone\_id="NCI\_CGAP\_GC6"  
/tissue\_type="pooled germ cell tumors"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Plasmid DNA from the normalized library  
was PCR-amplified and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive cDNA library (clones 1257096-1258631,  
from the same library (clones 1257096-1258631,  
1469064-1470983, and 1475592-1476743). Subtraction by  
Bento Soares and M. Fatima Bonaldo."

BASE COUNT 135 a 107 c 115 g 144 t

Query Match 6.1%; Score 42.2; DB 49; Length 501;  
Best Local Similarity 93.6%; Pred. No. 0.87; 3; Indels 0; Gaps 0;  
Matches 44; Conservative 0; Mismatches 0;

QY 536 TTCACATGCTAATCCAGCATTTGTGATAGCTGCAAACTCGGATAT 582

```

Db      256 TTCCATTGCTATCCAGCATTTGCGGATAGCTCCAAAGCGGATAT 210
      |||
RESULT 12
ID      HSM003697 standard; RNA; EST; 521 BP.
XX      AL039221;
XX      AL039221.1
XX      e1397556
XX      12-MAR-1999 (Rel. 59, Created)
XX      12-MAR-1999 (Rel. 59, Last updated, Version 1)
DE      Homo sapiens mRNA; EST: DKFZp(null),E131_f1 (from clone
DE      DKFZp(null),E131)
      EST: expressed sequence tag.
..4
OS      Homo sapiens (human)
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
OC      Primates; Catarrhini; Homnidae; Homo.
XX      [1]
XX      Poustka A., Klein M., Mewes W., Gassenhuber J., Wiemann S.;
XX      Submitted (12-MAR-1999) to the EMBL/Genbank/DBJ databases.
XX      M15, Am Klopferstritz 18a D-82152 Martinsried, GERMANY
RL      CLONE from S. Wiemann, sequenced by DKFZ within the CDNA
CC      sequencing consortium of the German Genome Project
CC      No s1 sequence available
CC      This clone is available at the RZPD in Berlin
CC      Please contact the RZPD, Ressourcenzentrum, Heubnerweg 6, 14059
CC      Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX      Key      Location/Qualifiers
FH      source      1. 521
FH      /db_xref="taxon:9606"
FH      /organism="Homo sapiens"
FH      /clone="DKFZp(null),E131"
FH      /clone_11b="(null) (synonym: hmcfl). Vector pSport1; host
FH      DH10B; sites NotI + SalI"
FH      /dev_stage="adult"
FH      /tissue_type="breast cancer"
XX      Sequence 521 BP; 62 A; 193 C; 176 G; 88 T; 2 other;
SQ
Query Match      6.1%; Score 41.8; DB 53; Length 521;
Best Local Similarity 47.4%; Pred. No. 1.1;
Matches 156; Conservative 0; Mismatches 169; Indels 4; Gaps 1;

```

```

QY      293 CGAGGTAGAGGGAAGATGTCATGCTCAGGCTTCAAGCCAAACCGAAGTACTGCGCTC 352
      |||
Db      393 GCCAGTGCCTGGGACCCCGGGGCTTCAGAGGAGGACCGAAGTCCGAGGAGCC 452
QY      353 TATCTTCAGAGGAGACCGAGCCACAGC 381
      |||
Db      453 CGAGTCCCGGACGAGGAGGCGTGCAGC 481
      |||
RESULT 13
AA697326      494 bp      mRNA      EST      28-NOV-1998
LOCUS      AA697326
DEFINITION      H102253.Sprine HL Drosophila melanogaster head Bluescript
Drosophila melanogaster cDNA clone H102253 Sprime similar to
M20316: D.melanogaster acetylcholine receptor-related protein mRNA,
complete cds, mRNA sequence.
ACCESSION      AA697326
NID      92700255
VERSION      AA697326.1 GI:2700255
KEYWORDS      EST.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 494)
AUTHORS      Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,
      Brokstein, P., Lewis, S. and Rubin, G.M.
      BDBP/HHMI Drosophila EST Project
      Unpublished (1997)
      On Nov 20, 1997 this sequence version replaced gi:1172316.
TITLE      JOURNAL
COMMENT      Contact: Harvey, D.
      G. M. Rubin-Molecular and Cell Biology
      University of California Berkeley
      539 USA, Berkeley, CA 94720-3200, USA
      Fax: 510 643 9947
      Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
      Plate: 22 row: E column: 5
      High quality sequence stop: 416.
FEATURES
      source      1. 494
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
      /clone="H102253"
      /clone_11b="HL Drosophila melanogaster head Bluescript"
      /sex="male and female"
      /dev_stage="adult"
      /lab_host="SOLR"
      /note="Organ: head-brain & sensory organ; Vector:
      Bluescript SK; Site: 1; EcorI; Site 2: XhoI; Constructed
      using Stratagene ZAP-cDNA Synthesis Kit. Oligo dT-primed
      and directionally cloned at EcoRI and XhoI in Bluescript
      SK(+/-)"
BASE COUNT      124 a 114 c 127 g 129 t
ORIGIN
Query Match      5.9%; Score 40.8; DB 37; Length 494;
Best Local Similarity 59.5%; Pred. No. 1.9;
Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```

A1492967  
 LOCUS 439 bp mRNA EST 18-MAR-1999  
 DEFINITION g247c02.x1 NCI CGAP\_Kid11 Homo sapiens CDNA clone IMAGE:2030018 3' similar to SW:CDNC\_HUMAN P49918 CYCLIN-DEPENDENT KINASE INHIBITOR 1C; contains TARI.12 TARI repetitive element ;, mRNA sequence.  
 ACCESSION A1492967  
 NID 94393970  
 VERSION A1492967.1 GI:4393970  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1. (bases 1 to 439)  
 NCICGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 TITLE Unpublished (1997)  
 JOURNAL On May 18, 1998 this sequence version replaced gi:3136536.  
 COMMENT

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/dbip/image/image.html

Insert Length: 908 Std Error: 0.00  
 Seq primer: 40UP from GIBCO  
 High quality sequence stop: 65.  
 Location/Qualifiers

FEATURES  
 source 1. 439  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="21922.1-3"  
 /clone="IMAGE:2030018"  
 /clone\_lib="NCI-CGAP\_Kid11"  
 /lab\_host="DH10B"  
 /note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP\_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 132376-132391, 145607-145675, and 150952-150985). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 60 a 145 c 187 g 47 t  
 ORIGIN

Query Match 5.7%; Score 39.4; DB 47; Length 439;  
 Best Local Similarity 53.3%; Pred. No. 4;  
 Matches 105; Conservative 0; Mismatches 91; Indels 1; Gaps 1;  
 Oy 138 GTCCTGCTGCCCCCTTCCATTTCCAGCCGCGCTCCACAGAGGTCACGGCGCGGGA 197  
 Db 46 GCGCGCCCTGCTGCTGCTGCGCGCTGCGCGCTGCTGAGCGCGCGCGCGG 105  
 Oy 198 GAGGTGAGCCGCGAGAGCTCGCGCGCGCGCGCGCTGCTGCGCGCGCAAGACGGC 257  
 Db 106 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT-GCTGCGCGCGCGCGCGC 164  
 Oy 258 GCTGGGAGTGGCTCTTTTCCGCGCGCGCGCGCGAGGTGAGGAGATGTCATG 317  
 Db 165 ACTGAGCGCGCGCGCGAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGAGACCGAG 224

Oy 318 TCAGGTTCAAGCCCA 334  
 Db 225 ACCGCGACCGAGCGCA 241  
 RESULT 15  
 A1192683  
 LOCUS 557 bp mRNA EST 29-OCT-1998  
 DEFINITION g62607.x1 Soares\_fetal\_lung\_NbHL19W Homo sapiens CDNA clone IMAGE:1743565 3' similar to SW:CDNC\_HUMAN P49918 CYCLIN-DEPENDENT KINASE INHIBITOR 1C; contains MER22.b3 TARI repetitive element ;, mRNA sequence.  
 ACCESSION A1192683  
 NID 93743892  
 VERSION A1192683.1 GI:3743892  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1. (bases 1 to 557)  
 NCICGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 TITLE Unpublished (1997)  
 JOURNAL On Jan 14, 1998 this sequence version replaced gi:1797633.  
 COMMENT

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 862 Std Error: 0.00  
 Seq primer: 40UP from GIBCO  
 High quality sequence stop: 118.  
 Location/Qualifiers

FEATURES  
 source 1. 557  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1743565"  
 /clone\_lib="Soares\_fetal\_lung\_NbHL19W"  
 /dev\_stage="19 weeks"  
 /lab\_host="DH10B (ampicillin resistant)"

/note="Organ: lung; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTACCAATCTGAGAGCGAGCGCGCGCAATTTTCTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. This library was constructed from the same fetus as the fetal heart library, Soares fetal heart NbHL19W."

BASE COUNT 62 a 193 c 242 g 60 t  
 ORIGIN

Query Match 5.7%; Score 39; DB 43; Length 557;  
 Best Local Similarity 53.3%; Pred. No. 5.4;  
 Matches 104; Conservative 0; Mismatches 90; Indels 1; Gaps 1;  
 Oy 138 GTCCTGCTGCCCCCTTCCATTTCCAGCCGCGCTCCACAGAGGTCACGGCGCGGGA 197  
 Db 68 GCGCGCCCTGCTGCTGCTGCGCGCTGCTGAGCGCGCGCGCTCGCGG 127  
 Oy 198 GAGGTGAGCCGCGAGAGCTCGCGCGCGCGCGCGCTGCTGCGCGCGCAAGACGGC 257  
 Db 128 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT-GGCTGCGCGCGCGCGC 186  
 Oy 258 GCTGGGAGTGGCTCTTTTCCGCGCGCGCGCGAGGTGAGGAGATGTCATG 317

Db 187 ACTGAGCCGGGGCCGAGCCGAGCCGGGGGGGGGGCCAGGAGCCGG 246  
OY 318 TCAGGGTTCAAGCC 332  
Db 247 ACCGCGAGCCGGAGCC 261

Search completed: September 17, 1999, 21:28:33  
Job time: 14297 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:04:21 ; Search time 1962.17 Seconds

(without alignments)  
59.970 Million cell updates/sec

Title: US-08-956-518a-100

Perfect score: 37

Sequence: 1 ACCCACACAGGCATATTCAGAGTTCCTGCTACATCG 37

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:\*  
1: gb\_bal:\*  
2: gb\_ba2:\*  
3: gb\_cm:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_p11:\*  
8: gb\_p12:\*  
9: gb\_p11:\*  
10: gb\_p12:\*  
11: gb\_p13:\*  
12: gb\_p14:\*  
13: gb\_p15:\*  
14: gb\_p16:\*  
15: gb\_p17:\*  
16: gb\_p18:\*  
17: gb\_p19:\*  
18: gb\_p20:\*  
19: gb\_p21:\*  
20: gb\_p22:\*  
21: gb\_p23:\*  
22: gb\_p24:\*  
23: gb\_p25:\*  
24: gb\_p26:\*  
25: gb\_p27:\*  
26: gb\_p28:\*  
27: gb\_p29:\*  
28: gb\_p30:\*  
29: gb\_p31:\*  
30: gb\_p32:\*  
31: gb\_p33:\*  
32: gb\_p34:\*  
33: gb\_p35:\*  
34: gb\_p36:\*  
35: gb\_p37:\*  
36: gb\_p38:\*  
37: gb\_p39:\*  
38: gb\_p40:\*  
39: gb\_p41:\*  
40: gb\_p42:\*  
41: gb\_p43:\*  
42: gb\_p44:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match Length	ID	Description
1	30	81.1	1559	9 HUMA7NAR

Result	Score	Match Length	ID	Description
1	30	81.1	1559	9 HUMA7NAR

## ALIGNMENTS

RESULT 1  
LOCUS HUMA7NAR 1559 bp mRNA PRI 18-AUG-1994  
DEFINITION Human a7 nicotinic acetylcholine receptor mRNA.  
ACCESSION L25827  
NID 9438616  
VERSION L25827.1 GI:438616  
KEYWORDS nicotinic acetylcholine receptor alpha 7 subunit; transmembrane protein.  
SOURCE Homo sapiens (library: Clontech HL1065b; ATCC 37433) fetus brain  
CDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (sites)  
AUTHORS Dovelette-Stamm, L., Monteggia, L.M., Donnelly-Roberts, D., Wang, M.T., Lee, J., Tian, J., and Giordano, T.  
TITLE Cloning and sequence of the human a7 nicotinic acetylcholine receptor  
JOURNAL Res. (1993) in press  
FEATURES  
Location/Qualifiers  
1..1559  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"



```

/dev_stage="fetus"
/tissue_type="brain"
/tissue_id="clontech HLI065b; ATCC 37433"
BASE COUNT      314 a      463 c      440 g      342 t
ORIGIN

Query Match      81.1%; Score 30; DB 9; Length 1559;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 CAGGCATATTCAGAGTTCCTGCTACATCG 37
      |||||||
Db      405 CAGGCATATTCAGAGTTCCTGCTACATCG 434

RESULT 2
      A7A      2087 bp      mRNA      PRI      01-JUN-1994
DEFINITION H.sapiens mRNA for neuronal nicotinic acetylcholine receptor
            alpha-7 subunit.
ACCESSION X70297
NID       9496606
VERSION   X70297.1 GI:496606
KEYWORDS  neuronal nicotinic acetylcholine receptor alpha-7 subunit.
SOURCE    human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2087)
AUTHORS   Peng, X., Katz, M., Gerzanich, V., Anand, R. and Lindstrom, J.
TITLE     Human alpha 7 acetylcholine receptor: cloning of the alpha 7
            subunit from the SH-SY5Y cell line and determination of
            pharmacological properties of native receptors and functional alpha
            7 homomers expressed in Xenopus oocytes
            Mol. Pharmacol. 45 (3), 546-554 (1994)
JOURNAL   94195283
MEDLINE   2 (bases 1 to 2087)
AUTHORS   Katz, M.
TITLE     Direct Submission
JOURNAL   Submitted (04-FEB-1993) M. Katz, Univ of Pennsylvania School of
            Medicine, Dept of Neuroscience, 365h & Hamilton Walk, Rm 235
            Stemmler Hall, Philadelphia, PA 19104, USA
FEATURES
    source
        1..2087
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /tissue_type="brain"
            /cell_type="neuroblastoma"
            /cell_line="SH-SY5Y"
            /clone_id="lambda zap II"
            /clone="SHSY3"
            104..175
                /product="neuronal nicotinic acetylcholine receptor
                alpha-7 subunit"
            104..1612
                /codon_start=1
                /product="neuronal nicotinic acetylcholine receptor
                alpha-7 subunit"
                /protein_id="CAA49778.1"
                /db_xref="PID:9496607"
                /db_xref="GI:496607"
            /db_xref="SWISS-PROT:P36544"
            /translation="MRCSPGVWGLAASLLHVSLOGEFORLKYELKYNPLERPV
            ANDSPLVYFSLQIMDYDEKNOVLITNIMLOMSWTDHLYOMNVSEYGVKIVRF
            PDGQIKRPDLILNSADERFATFTNVLVNSGCOYLPGLFKSSCYIDVWPEFD
            VOCHKIKGSMVSGWGLDLOMOEADISGTYIPNGEMDVGIPGKRSERFECCKEYF
            DVTEFTVTRRTIYGLNLIPCVLISALALVFLPADSGEKISGIVLISLVLEM
            LVAETMPATSDSVPLAIFYASTIMVIGLSVAVTVIADYHHDDPGGKMPKTRVI
            LLMNCAMFLRMKRPEDKVRPAQKORRCSLASVEMSAVAPPASNGVLIYIGFGL
            DGVAQVTPDSGVVCGRMACSPTHDEHLHGCGPPEDGPDLAILEEVYIANRFKQ
            DESRAVCSWEMKFAVCYVRLCLMAFSVFTICTIGILMSAPNVEAVSKDFA"
            176..1609

```

mat\_peptide

```

BASE COUNT      441 a      601 c      573 g      472 t
ORIGIN

Query Match      81.1%; Score 30; DB 10; Length 2087;
Best Local Similarity 100.0%; Pred. No. 0.0052;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 CAGGCATATTCAGAGTTCCTGCTACATCG 37
      |||||||
Db      531 CAGGCATATTCAGAGTTCCTGCTACATCG 560

RESULT 3
      HSNACHRA7      1509 bp      mRNA      PRI      22-JAN-1998
LOCUS      H.sapiens mRNA for nicotinic acetylcholine receptor alpha7 subunit
DEFINITION precursor.
ACCESSION Y08420
NID       92808623
VERSION   Y08420.1 GI:2808623
KEYWORDS  nAChR gene; nicotinic acetylcholine receptor alpha 7 subunit.
SOURCE    human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1509)
AUTHORS   Groot Kormelink, P.J. and Luyten, W.H.
TITLE     Cloning and sequence of full-length cDNAs encoding the human
            neuronal nicotinic acetylcholine receptor (nAChR) subunits betas
            and beta4 and expression of seven nAChR subunits in the human
            neuroblastoma cell line SH-SY5Y and/or IMR-32
            FEBS Lett. 400 (3), 309-314 (1997)
JOURNAL   97162233
MEDLINE   2 (bases 1 to 1509)
AUTHORS   Groot Kormelink, P.J.
TITLE     Direct Submission
JOURNAL   Submitted (27-SEP-1996) P.J. Groot Kormelink, Janssen Research
            Foundation, Exp. Mol. Biol. Dept., Turnhoutseweg 30, B-2240 Beerse,
            Belgium
REMARK    Revised by author 22-JAN-1998
COMMENT   On Jan 25, 1998 this sequence version replaced gi:1702915.
FEATURES
    source
        1..1509
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /cell_type="neuroblastoma"
            /cell_line="IMR-32"
            /cell_line="SH-SY5Y"
            1..66
                /gene="nAChRA7"
            1..1509
                /gene="nAChRA7"
                /codon_start=1
                /product="nicotinic acetylcholine receptor alpha7 subunit
                precursor"
                /protein_id="CAA69697.1"
                /db_xref="PID:e1246208"
                /db_xref="PID:g2808624"
                /db_xref="GI:2808624"
            /translation="MRCSPGVWGLAASLLHVSLOGEFORLKYELKYNPLERPV
            ANDSPLVYFSLQIMDYDEKNOVLITNIMLOMSWTDHLYOMNVSEYGVKIVRF
            PDGQIKRPDLILNSADERFATFTNVLVNSGCOYLPGLFKSSCYIDVWPEFD
            VOCHKIKGSMVSGWGLDLOMOEADISGTYIPNGEMDVGIPGKRSERFECCKEYF
            DVTEFTVTRRTIYGLNLIPCVLISALALVFLPADSGEKISGIVLISLVLEM
            LVAETMPATSDSVPLAIFYASTIMVIGLSVAVTVIADYHHDDPGGKMPKTRVI
            LLMNCAMFLRMKRPEDKVRPAQKORRCSLASVEMSAVAPPASNGVLIYIGFGL
            DGVAQVTPDSGVVCGRMACSPTHDEHLHGCGPPEDGPDLAILEEVYIANRFKQ
            DESRAVCSWEMKFAVCYVRLCLMAFSVFTICTIGILMSAPNVEAVSKDFA"
            1..1509
            /gene="nAChRA7"
            /gene="nAChRA7"
            67..1506
            /gene="nAChRA7"

```

gene  
mat\_peptide

BASE COUNT 297 a 451 c 429 g 332 t  
ORIGIN

Query Match 81.1%; Score 30; DB 10; Length 1509;  
Best Local Similarity 100.0%; Pred. No. 0.0051;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 CAGCATATTCAGAGTCTCCTACATCG 37  
|||||  
Db 428 CAGCATATTCAGAGTCTCCTACATCG 457

RESULT 4  
HSU40583 1977 bp mRNA PRI 19-DEC-1995  
LOCUS Human alpha 7 neuronal nicotinic acetylcholine receptor mRNA,  
DEFINITION complete cds.  
ACCESSION U40583  
GI:1125076  
VERSION 1  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1977)  
Logel, J., Drebing, C., Barnhart, M., Antle, C. and Leonard, S.  
Nucleotide Sequence and Transcript Size of the Alpha-7 Neuronal  
Nicotinic Acetylcholine Receptor in Human Postmortem Brain  
Unpublished  
2 (bases 1 to 1977)  
Leonard, S.  
Direct Submission  
Submitted (13-NOV-1995) Sherry Leonard, University of Colorado  
Health Sciences Center, C-268-71 Pharmacology, 4200 E. Ninth Ave,  
Denver, CO 80262, USA  
FEATURES  
source Location/Qualifiers  
1. 1977  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="PBshalpha7"  
/clone\_lib="Clontech lambda gt10 cDNA library: human  
hippocampus"  
/sex="male"  
/tissue\_type="hippocampus"  
/dev\_stage="20 yr old adult"  
1. 7  
8. 73  
8. 1516  
/codon\_start=1  
/product="alpha 7 neuronal nicotinic acetylcholine  
receptor"  
/protein\_id="AA83561.1"  
/db\_xref="PID:g1125077"  
/db\_xref="GI:1125077"  
/translation="MRCSPGVWLAASLHVSLOGEFORKLYKELYNPNLEPV  
ANDSPLTVFSLNLIQIMVDERKNVLTINIMLOMSWTIDYLIOMNVESEPGVTVF  
PDGQIMKPDILVNSADERDAPHTNVLVNSGHCYLPDGLFSSCYIDVRFPD  
VORCKLFGSMISGWSLDLOMGADISGYPNGEMDLVGIPIKRSERFECEKEYP  
DVFETVMRRRTLYGILNLIIPCLISALALFLPADSGEKISLIGITVLSTLVN  
LVAELMIPASDSVPLAQYFASMTIIVGLSVVTVVLYOHHDHDPGGKMPKTRVY  
LLMNCWFLMKRPGEDKVRACQKORCSLASVENSAYAPPAASGNLLYIGFRGL  
DGWACVPTPSGVVCGMACSPHDEHLLHGSGPPEDDPLAKTLEVRVIANFRQ  
DESEAVSEMKFAACVDRCLMAFSVFTIICIGILMSAPNFEAVSKDFA"  
1514. 1977  
3' UTR  
polyA-site  
BASE COUNT 426 a 567 c 524 g 460 t  
ORIGIN

Query Match 81.1%; Score 30; DB 10; Length 1977;  
Best Local Similarity 100.0%; Pred. No. 0.0052;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 CAGCATATTCAGAGTCTCCTACATCG 37  
|||||  
Db 435 CAGCATATTCAGAGTCTCCTACATCG 464

RESULT 5  
HSU62436 1876 bp mRNA PRI 11-JAN-1997  
LOCUS Human nicotinic acetylcholine receptor alpha7 subunit precursor,  
DEFINITION mRNA, complete cds.  
ACCESSION U62436  
GI:1458119  
VERSION 1  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1876)  
Elliot, K.J., Ellis, S.B., Berckhan, K.J., Urrutia, A.,  
Chavez-Noriega, L.E., Johnson, E.C., Velicelebi, G. and Harpold, M.M.  
Comparative structure of human neuronal alpha 2-alpha 7 and beta  
2-beta 4 nicotinic acetylcholine receptor subunits and functional  
expression of the alpha 2, alpha 3, alpha 4, alpha 7, beta 2, and  
beta 4 subunits  
J. Mol. Neurosci. 7 (3), 217-228 (1996)  
97062879  
2 (bases 1 to 1876)  
Elliot, K.J.  
Direct Submission  
Submitted (28-JUN-1996) Kathryn J. Elliot, SIBIA Neurosciences,  
Inc., 505 Coast Blvd. So., La Jolla, CA 92037, USA  
FEATURES  
source Location/Qualifiers  
1. 1876  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="K6alpha7.3"  
/clone\_lib="SIBIA lambda gt11 library #3 (M. Williams)"  
/cell\_line="IMR32"  
/cell\_type="neuroblastoma"  
1. 72  
73. 138  
73. 1581  
/codon\_start=1  
/product="nicotinic acetylcholine receptor alpha7 subunit  
precursor"  
/protein\_id="AAB40114.1"  
/db\_xref="PID:g1458120"  
/db\_xref="GI:1458120"  
/translation="MRCSPGVWLAASLHVSLOGEFORKLYKELYNPNLEPV  
ANDSPLTVFSLNLIQIMVDERKNVLTINIMLOMSWTIDYLIOMNVESEPGVTVF  
PDGQIMKPDILVNSADERDAPHTNVLVNSGHCYLPDGLFSSCYIDVRFPD  
VORCKLFGSMISGWSLDLOMGADISGYPNGEMDLVGIPIKRSERFECEKEYP  
DVFETVMRRRTLYGILNLIIPCLISALALFLPADSGEKISLIGITVLSTLVN  
LVAELMIPASDSVPLAQYFASMTIIVGLSVVTVVLYOHHDHDPGGKMPKTRVY  
LLMNCWFLMKRPGEDKVRACQKORCSLASVENSAYAPPAASGNLLYIGFRGL  
DGWACVPTPSGVVCGMACSPHDEHLLHGSGPPEDDPLAKTLEVRVIANFRQ  
DESEAVSEMKFAACVDRCLMAFSVFTIICIGILMSAPNFEAVSKDFA"  
139. 1578  
/product="nicotinic acetylcholine receptor alpha7 subunit"  
1582. 1876  
3' UTR  
mat-peptide  
BASE COUNT 369 a 553 c 531 g 423 t  
ORIGIN

Query Match 81.1%; Score 30; DB 10; Length 1876;  
Best Local Similarity 100.0%; Pred. No. 0.0052;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 8 CAGCATATTCAGAGTCTCCTACATCG 37  
|||||

Db 500 CAGGCATATTCAGAGTTCCTGCTACATCG 529

RESULT 6  
AF029838 641 bp mRNA PRI 16-DEC-1998  
LOCUS Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.  
DEFINITION AF029838  
ACCESSION AF029838  
NID 93757793  
VERSION AF029838.1 GI:3757793  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
AUTHORS 1 (bases 1 to 641)  
Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J.,  
Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K.,  
Bulting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.  
Genomic organization and partial duplication of the human alpha7  
neuronal nicotinic acetylcholine receptor gene  
Genomics 52 (2), 173-185 (1998)

TITLE  
JOURNAL 9900837  
MEDLINE  
REFERENCE  
AUTHORS 2 (bases 1 to 641)  
Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,  
Breese, C., Davis, A., Hopkins, J. and Freedman, R.  
Direct Submission  
Submitted (15-OCT-1997) Psychiatry, University of Colorado Health  
Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA  
Location/Qualifiers

FEATURES  
source  
1..641  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15q14"  
/clone="CHRNA7-DR1"  
/note="RACE product A/B/C/D"  
<1..>641  
/note="alpha 7 neuronal nicotinic receptor mRNA; used in  
alternative splicing"  
192 c 163 g 135 t

misc.feature  
BASE COUNT 151 a 192 c 163 g 135 t  
ORIGIN

Query Match 81.1%; Score 30; DB 11; Length 641;  
Best Local Similarity 100.0%; Pred. No. 0.0048; Mismatches 0; Indels 0; Gaps 0;  
atches 30; Conservative

QY 8 CAGGCATATTCAGAGTTCCTGCTACATCG 37  
|||||  
Db 612 CAGGCATATTCAGAGTTCCTGCTACATCG 641

RESULT 7  
AF029839 689 bp mRNA PRI 16-DEC-1998  
LOCUS Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.  
DEFINITION AF029839  
ACCESSION AF029839  
NID 93757794  
VERSION AF029839.1 GI:3757794  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
AUTHORS 1 (bases 1 to 689)  
Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J.,  
Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K.,  
Bulting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.  
Genomic organization and partial duplication of the human alpha7  
neuronal nicotinic acetylcholine receptor gene  
Genomics 52 (2), 173-185 (1998)

TITLE  
JOURNAL 9900837  
MEDLINE  
REFERENCE  
AUTHORS 2 (bases 1 to 689)  
Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,  
Breese, C., Davis, A., Hopkins, J. and Freedman, R.  
Direct Submission  
Submitted (15-OCT-1997) Psychiatry, University of Colorado Health  
Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA  
Location/Qualifiers

FEATURES  
source  
1..689  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15q14"  
/clone="CHRNA7-DR2"  
/note="RACE product A/C/D"  
<1..>689  
/note="alpha 7 neuronal nicotinic receptor mRNA; used in  
alternative splicing"  
210 c 196 g 141 t

misc.feature  
BASE COUNT 142 a 210 c 196 g 141 t  
ORIGIN

Query Match 81.1%; Score 30; DB 11; Length 689;  
Best Local Similarity 100.0%; Pred. No. 0.0049; Mismatches 0; Indels 0; Gaps 0;  
Matches 30; Conservative

QY 8 CAGGCATATTCAGAGTTCCTGCTACATCG 37  
|||||  
Db 660 CAGGCATATTCAGAGTTCCTGCTACATCG 689

RESULT 8  
AF036903 1712 bp mRNA PRI 16-OCT-1998  
LOCUS Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor  
DEFINITION AF036903  
ACCESSION AF036903  
NID 93757807  
VERSION AF036903.1 GI:3757807  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
AUTHORS 1 (bases 1 to 1712)  
Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J.,  
Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K.,  
Bulting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.  
Genomic organization and partial duplication of the human alpha7  
neuronal nicotinic acetylcholine receptor gene (CHRNA7)  
Genomics 52 (2), 173-185 (1998)

TITLE  
JOURNAL 9900837  
MEDLINE  
REFERENCE  
AUTHORS 2 (bases 1 to 1712)  
Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,  
Breese, C., Davis, A., Hopkins, J. and Freedman, R.  
Direct Submission  
Submitted (03-DEC-1997) Psychiatry, University of Colorado Health  
Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA  
Location/Qualifiers

FEATURES  
source  
1..1712  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/db\_xref="dbEST:W52861"  
/db\_xref="dbEST:AA037389"  
/chromosome="15"  
/clone="1321242"  
/clone="11D"=Soares papathyroid tumor NHPA library,  
Research Genetics/IMAGE Consortium, LBNL  
<1..1712  
/note="alpha-7 neuronal nicotinic acetylcholine receptor;  
alternatively spliced"  
489 c 457 g 410 t

misc.feature  
BASE COUNT 356 a 489 c 457 g 410 t  
ORIGIN

REFERENCE  
AUTHORS 2 (bases 1 to 689)  
Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,  
Breese, C., Davis, A., Hopkins, J. and Freedman, R.  
Direct Submission  
Submitted (15-OCT-1997) Psychiatry, University of Colorado Health  
Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA  
Location/Qualifiers

FEATURES  
source  
1..689  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15q14"  
/clone="CHRNA7-DR2"  
/note="RACE product A/C/D"  
<1..>689  
/note="alpha 7 neuronal nicotinic receptor mRNA; used in  
alternative splicing"  
210 c 196 g 141 t

misc.feature  
BASE COUNT 142 a 210 c 196 g 141 t  
ORIGIN

Query Match 81.1%; Score 30; DB 11; Length 689;  
Best Local Similarity 100.0%; Pred. No. 0.0049; Mismatches 0; Indels 0; Gaps 0;  
Matches 30; Conservative

QY 8 CAGGCATATTCAGAGTTCCTGCTACATCG 37  
|||||  
Db 660 CAGGCATATTCAGAGTTCCTGCTACATCG 689

RESULT 8  
AF036903 1712 bp mRNA PRI 16-OCT-1998  
LOCUS Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor  
DEFINITION AF036903  
ACCESSION AF036903  
NID 93757807  
VERSION AF036903.1 GI:3757807  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
AUTHORS 1 (bases 1 to 1712)  
Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J.,  
Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K.,  
Bulting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.  
Genomic organization and partial duplication of the human alpha7  
neuronal nicotinic acetylcholine receptor gene (CHRNA7)  
Genomics 52 (2), 173-185 (1998)

TITLE  
JOURNAL 9900837  
MEDLINE  
REFERENCE  
AUTHORS 2 (bases 1 to 1712)  
Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,  
Breese, C., Davis, A., Hopkins, J. and Freedman, R.  
Direct Submission  
Submitted (03-DEC-1997) Psychiatry, University of Colorado Health  
Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA  
Location/Qualifiers

FEATURES  
source  
1..1712  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/db\_xref="dbEST:W52861"  
/db\_xref="dbEST:AA037389"  
/chromosome="15"  
/clone="1321242"  
/clone="11D"=Soares papathyroid tumor NHPA library,  
Research Genetics/IMAGE Consortium, LBNL  
<1..1712  
/note="alpha-7 neuronal nicotinic acetylcholine receptor;  
alternatively spliced"  
489 c 457 g 410 t

misc.feature  
BASE COUNT 356 a 489 c 457 g 410 t  
ORIGIN

Query Match  
Best Local Similarity 81.1%; Score 30; DB 11; Length 1712;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGGCATATTCAGAGTCTCGCTACATCG 37  
DB 172 CAGGCATATTCAGAGTCTCGCTACATCG 201  
|||||

RESULT 9  
LOCUS AF037646 1836 bp RNA PRI 16-OCT-1998  
DEFINITION Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor  
ACCESSION AF037646  
VERSION AF037646.1 GI:3757808  
KEYWORDS human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 1836)  
Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J.,  
Moore, T., Jacobs, S., Meriwether, J., Choi, M., Kim, E., Walton, K.,  
Bulling, K., Davis, A., Breese, C., Freedman, R., and Leonard, S.  
Genomic organization and partial duplication of the human alpha7  
neuronal nicotinic acetylcholine receptor gene (CHRNA7)  
Genomics 52 (2), 173-185 (1998)  
99000837  
2 (bases 1 to 1836)  
Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,  
Breese, C., Davis, A., Hopkins, J., and Freedman, R.  
Direct Submission  
Submitted (03-DEC-1997) Psychiatry, University of Colorado Health  
Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA  
FEATURES  
Source  
1. 1836  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/db\_xref="dbest:W03952"  
/db\_xref="dbest:N73891"  
/chromosome="15"  
/clone="297134"  
/clone\_1b="Scars fetal liver spleen library, Research  
Genetics/IMAGE Consortium, LNL"  
/note="unprocessed mRNA with intron"  
<1. 1836  
/note="alpha-7 neuronal nicotinic acetylcholine receptor  
precursor; intron included in 5' region"

BASE COUNT 383 a 503 c 469 g 481 t

misc-feature

Query Match  
Best Local Similarity 100.0%; Pred. No. 0.0052;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGGCATATTCAGAGTCTCGCTACATCG 37  
DB 294 CAGGCATATTCAGAGTCTCGCTACATCG 323  
|||||

RESULT 10  
LOCUS MUSNARS 1848 bp mRNA ROD 02-FEB-1999  
DEFINITION Mus musculus neuronal nicotinic acetylcholine receptor subunit  
alpha 7 mRNA, complete cds.  
ACCESSION L37663  
VERSION L37663.1 GI:790853  
KEYWORDS neuronal acetylcholine receptor subunit alpha 7 unit.

SOURCE Mus musculus (strain BALB/c, sub-species domesticus) neonatal brain  
CDNA to mRNA.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1848)  
Orr-Uttreger, A., Seldin, M.F., Baldini, A., and Beaudet, A.L.  
Cloning and mapping of the mouse alpha 7-neuronal nicotinic  
acetylcholine receptor  
Genomics 26 (2), 399-402 (1995)  
95324936  
FEATURES  
source  
Location/Qualifiers  
1. 1848  
/organism="Mus musculus"  
/strain="BALB/c"  
/sub-species="domesticus"  
/db\_xref="taxon:10090"  
/dev\_stage="neonatal"  
/tissue\_type="brain"  
1. 50  
51. 1559  
/codon\_start=1  
/product="neuronal acetylcholine receptor subunit alpha 7"  
/protein\_id="AAC42053.1"  
/db\_xref="PID:9790854"  
/db\_xref="GI:790854"  
/translation="MCGRRGIMVLAALHVSLOEFPORLYKELVKNYPLERPV  
PDQIWMFDILVNSADREDAFTNIVLVASGCOYLPGLKSCYIDVREPPD  
VOCKLKFSGMSYGGMSLDLQMEADISYIPNGEMDMIGPKRNRKFECCKEP  
DVTYTVRRRTLYGGLIPCVLIALLVFLPADSGEKSLDITVLSITVM  
LVNAWFLPMSDVPILTAOYFASMTIVGLSVYTVYLVTHHDPDGGKPKWTRII  
LNMCAWFLPMSDVPILTAOYFASMTIVGLSVYTVYLVTHHDPDGGKPKWTRII  
EGMHCAPTSPSGVYCGSLASPHDEHLMGTHPSDDPLAKILEEVIYANFRQ  
DESEVISEKRPACVYDRLCMAFVSFTICTIGILMSAPNFEVANSKQFA"  
1560. 1848

BASE COUNT 406 a 487 c 500 g 455 t

3' UTR

Query Match  
Best Local Similarity 77.3%; Score 28.6; DB 12; Length 1848;  
Matches 31; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CCACAGGCATATTCAGAGTCTCGCTACATCG 37  
DB 473 CCCTCAGGCATATTCAGAGTCTCGCTACATCG 507  
|||||

RESULT 11  
LOCUS HSCHRNA7A 1555 bp mRNA PRI 03-MAR-1994  
DEFINITION H. sapiens CHRNA7 mRNA, 3' end.  
ACCESSION Z23141  
VERSION Z23141.1 GI:457736  
KEYWORDS alpha7 nicotinic receptor subunit; cholinergic receptor; CHRNA7  
gene.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 1555)  
Chini, B., Raimond, E., Elgoyhen, A.B., Morali, D., Balzarotti, M. and  
Heinemann, S.  
Molecular cloning and chromosomal localization of the human alpha  
7-nicotinic receptor subunit gene (CHRNA7)  
Genomics 19 (2), 379-381 (1994)  
94245214  
2 (bases 1 to 1555)  
Chini, B.  
Direct Submission  
Submitted (29-JUN-1993) BICE CHINI, CCIPE, Rue de la Cardonille,

42. .1550  
/gene="nicotinic receptor alpha 7 subunit" Protein sequence  
CDS

gene 42..1350  
/gene="nicotinic receptor alpha 7 subunit-  
42..1550  
CDS 42..1550  
/gene="nicotinic receptor alpha 7 subunit"  
/note="this sequence comes from fig. 1. Protein sequence



Sun Sep 19 10:26:01 1999

us-08-956-518a-100.rge

Query Match 60.0%; Score 22.2; DB 10; Length 2251;  
Best local similarity 77.1%; Pred. No. 12;  
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 2 CCCACACAGGCATATTCACAGAGTTCCTGCTACATC 36  
||||| ||||||| ||| |||  
Db 1053 CCCACACAGGCATATTCACAAACACCTGCTCCATC 1019  
Search completed: September 17, 1999, 22:04:23  
Job time: 16446 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:34:34 ; Search time 425.19 Seconds

(without alignments)  
21.772 Million cell updates/sec

Title: US-08-956-518a-100

Perfect score: 37

Sequence: 1 ACCGACAGCATATTCAGAGTCTCTCATACG 37

Scoring table: IDENTITY\_NNC

Searched: 311585 seqs, 125096042 residues

Database: N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	30	81.1	1876	1	T48239	Neuronal nicotinic
2	30	81.1	1876	1	V12197	Human neuronal nic
3	30	81.1	1876	1	V44687	V2747 variant huma
4	22	59.5	2769	1	T59196	Neuronal alpha-bun
5	21	56.8	3496	1	T48237	Neuronal nicotinic
6	21	56.8	2082	1	T59527	Alpha4 subunit of
7	21	56.8	2082	1	T59528	Alpha4 subunit of
8	21	56.8	2082	1	V12196	Human neuronal nic
9	20.8	56.2	51952	1	V26084	Tomato pest resist
10	19.6	53.0	2277	1	Q90387	Alpha 2 subunit of
11	19.6	53.0	2664	1	T48235	Neuronal nicotinic
12	19.6	53.0	2277	1	V12199	Human neuronal nic
13	19.2	51.9	377	1	V28636	Human aspartic pro
14	19.2	51.9	1375	1	V28623	Human aspartic pro
15	19.2	51.9	452	1	V28626	Human aspartic pro
16	19.2	51.9	1298	1	V31665	Nucleotide sequenc
17	19.2	51.9	1645	1	V27037	Human napsin A gen
18	19.2	51.9	1353	1	V27036	Human napsin A CDN
19	19.2	51.9	1910	1	V27038	Human napsin B CDN
20	19.2	51.4	6273	1	V52154	Streptococcus pneu
21	18.8	50.8	812	1	T13155	Turkey prepro-VIP
22	18.8	50.8	2101	1	T59197	Neuronal alpha-bun
23	18.8	50.8	402	1	V28632	Human aspartic pro
24	18.8	50.8	1860	1	X37192	Recombinant PGO-11
25	18.8	50.8	741	1	X37189	Recombinant PGO-9C
26	18.8	50.8	1476	1	X37190	Recombinant PGO-9C
27	18.8	50.8	1125	1	X37191	Recombinant PGO-11
28	18.8	50.8	651	1	X37193	Recombinant PGO-11
29	18.8	50.8	1386	1	X37194	Recombinant PGO-8C
30	18.8	50.8	110	1	X37194	PCR primer seq. ID
31	18.4	49.7	1794	1	Q33018	Gene encoding the
32	18.4	49.7	377	1	Q61211	Human brain Expres
33	18.4	49.7	1429	1	V04237	Arabidopsis C-14 s
34	18.4	49.7	6588	1	V04238	Arabidopsis C-14 s
35	18.4	49.7	2065	1	V29062	BRCA1 modulator pr
36	18.4	49.7	11050	1	V52241	Streptococcus pneu
37	18.4	49.7	11050	1	V74407	R. communis syntha
38	18.2	49.2	1845	1	Q2614	Synthase factor B
39	18.2	49.2	1845	1	Q63896	Synthase factor B
40	18.2	49.2	1672	1	O63895	DUB-1 enhancer/pro
41	18.2	49.2	1559	1	T91855	Modified tissue pl
42	18.2	49.2	1575	1	N82185	Mouse receptor ME2
43	18	48.6	10195	1	T85320	

## ALIGNMENTS

C 44 18 48.6 275 1 V88255  
C 45 17.8 48.1 6567 1 Q03324

EST clone GE68. Ne  
Elmertia tenella ge

## RESULT 1

T48239  
ID T48239 standard; DNA; 1876 BP.

AC T48239;

DT 09-APR-1997 (first entry)

DE Neuronal nicotinic acetylcholine receptor alpha-7 subunit DNA.

KW Neuronal nicotinic acetylcholine receptor; nAChR; neurotransmitter;

OS Homo sapiens.

FH Key

FT Location/Qualifiers

PN CDS

PD WO9641876-A1.

PE 27-DEC-1996.

PF 07-JUN-1996; 009775.

PR 07-JUN-1995; US-484722.

PA (SIBI-) SIBIA NEUROSCIENCES INC.

PI Elliott KJ, Harpold MM;

DR WPI: 97-065463/06.

DR P-PSDB; W09025.

PT Nucleic acids encoding nicotinic acetylcholine receptor sub-units -

PS used in screening to determine the effect of drugs on the receptor

PT Disclosure; Page 71-73; 108bp; English.

CC A DNA sequence (T48239) codes for the alpha-7 subunit (W09025) of

the human neuronal nicotinic acetylcholine receptor (nAChR). Host

cells, esp. mammalian cells or amphibian oocytes, carrying alpha-7

nucleic acids, opt. in combination with other alpha and/or beta

CC subunit nucleic acids (see also T48232-38, T48240-41), express

recombinant nAChR subunits useful for identifying cpds. that

modulate the activity of human nAChRs.

CC Sequence 1876 BP; 369 A; 553 C; 531 G; 423 T;

SO

Query Match

Best Local Similarity 100.0%; Pred. No. 0.0016;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGGCATATTCAGAGTCTCTCATACG 37

DB 500 CAGGCATATTCAGAGTCTCTCATACG 529

RESULT 2

V12197

ID V12197 standard; CDNA; 1876 BP.

AC V12197;

DT 14-MAY-1998 (first entry)

DE Human neuronal nicotinic acetylcholine receptor alpha-7 subunit CDNA.

KW Human, neuronal nicotinic acetylcholine receptor; alpha-7 subunit;

OS Homo sapiens.

FH Key

FT Location/Qualifiers

PN CDS

PD WO9420617-A2.

PE 15-SEP-1994.

PF 08-MAR-1994; U02447.

PR 08-MAR-1993; US-028031.

PA (SALK-) SALK INST BIOTECHNOLOGY IND ASSOC.

PA (SIBI-) SIBIA NEUROSCIENCES INC.

PI Elliott KJ, Ellis SB, Harpold MM;

DR WPI: 94-303024/37.

DR P-PSDB; W44153.

PT Human neuronal nicotinic acetylcholine receptor subunits and DNA -



PT also transformed cells useful for screening cpds. which modulate  
 PT activity of the receptor  
 PS Claim 8; Page 78-79; 99pp: English.  
 CC The present sequence encodes a human neuronal nicotinic acetylcholine  
 CC receptor (nAChR) subunit. The cells expressing the alpha and/or beta  
 CC nAChR subunits may be used in a method of screening compounds to  
 CC identify any which modulate the activity of human neuronal nAChR.  
 CC Subunit specific antibodies may be used to monitor the distribution  
 CC and expression density of various subunits in normal vs diseased brain  
 CC tissues. Testing of single receptor subunits or specific receptor  
 CC subunit combinations with a variety of potential agonists or antagonists  
 CC provides information with respect to the function and activity of the  
 CC individual subunits and should lead to the identification and design of  
 CC compounds that are capable of very specific interaction with one or  
 CC more receptor subtypes. The resulting drugs should exhibit fewer  
 CC unwanted side effects than drugs identified e.g. screening with cells  
 CC that express a variety of subtypes  
 SQ Sequence 1876 BP; 369 A; 533 C; 530 G; 423 T;  
 81.1%; Score 30; DB 1; Length 1876;  
 est Local Similarity 100.0%; Pred. No. 0.0016;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 8 CAGGCATATTCAGAGTCTCTGCTACATCG 37  
 ||||||||||||||||||||||||||||  
 DB 500 CAGGCATATTCAGAGTCTCTGCTACATCG 529  
 RESULT 3  
 ID V44687 standard; CDNA; 1590 BP.  
 AC V44687;  
 DT 09-OCT-1998 (first entry) nAChR coding sequence.  
 DE V274T variant human alpha7 nAChR coding sequence; cancer;  
 KW Alpha7 nAChR; alpha7 nicotinic acetylcholine receptor subunit; therapy;  
 KW neurodegeneration; enzyme dysfunction; affective disorder; disease;  
 KW immune dysfunction; diabetic neuropathy; Alzheimer's disease;  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 9..1517  
 FT /tag- a  
 FT CDS  
 WT09828331-A2.  
 PN 02-JUL-1998.  
 PD 22-DEC-1997; U23405.  
 PR 20-DEC-1996; US-771737.  
 PR (ABBO) ABBOTT LAB.  
 BR Briggs CA, Gopalakrishnan M, McKenna DG, Monteggia LM,  
 Roch U, Sullivan JP, Touma E;  
 P-PSDB; W69216.  
 DR WPI: 98-377593/32.  
 DR Nucleic acid encoding variant of human alpha7 nicotinic  
 PT acetylcholine receptor subunit - used to identify modulators of  
 PT the receptor, potentially useful for treating neuro-degeneration,  
 PT cancer, affective disorders etc.  
 PS Claim 14; Fig 2; 44pp: English.  
 CC This sequence encodes the V274T variant of human alpha7 nicotinic  
 CC acetylcholine receptor (nAChR) subunit and to identify modulators of  
 CC the DNA are used to express the protein and to identify modulators of  
 CC alpha7 nAChR activity or cytoprotective agents, e.g. antisense  
 CC compounds or antagonists that are potentially useful for treating  
 CC neurodegeneration, enzyme dysfunction, affective disorders and immune  
 CC dysfunction, such as cancer, post-hepatic neuropathy, diabetic  
 CC neuropathy, osteoarthritis, Alzheimer's or Parkinson's diseases, kuru,  
 CC psychosis and schizophrenia. Probes based on the DNA are used to detect  
 CC the DNA in usual hybridisation or amplification tests, while monoclonal  
 CC antibodies are used to detect the protein for diagnosis (in vitro or by  
 CC in situ immuno-fluorescent assay). Compared with wild-type alpha7 nAChR,  
 CC the protein has about 100-fold greater sensitivity to cholinergic  
 CC receptor agonists (nicotine or acetylcholine) and response to these  
 CC agonists decays more slowly, but the wild-type inward rectification is  
 CC retained.

SQ Sequence 1590 BP; 322 A; 471 C; 449 G; 348 T;  
 81.1%; Score 30; DB 1; Length 1590;  
 Query Match  
 Best Local Similarity 100.0%; Pred. No. 0.0015;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 8 CAGGCATATTCAGAGTCTCTGCTACATCG 37  
 ||||||||||||||||||||||||||||  
 DB 436 CAGGCATATTCAGAGTCTCTGCTACATCG 465  
 RESULT 4  
 ID T59196 standard; CDNA; 2769 BP.  
 AC T59196; 17-JUN-1997 (first entry)  
 DT 17-JUN-1997  
 DE Neuronal alpha-bungarotoxin binding protein alpha1 subunit CDNA.  
 KW Neuronal alpha-bungarotoxin binding protein alpha 1; cholinergic;  
 KW ligand binding; ion channel; ss.  
 OS Gallus sp.  
 FH Key Location/Qualifiers  
 FT cds 71..1513  
 FT /tag- a  
 FT signal\_peptide 71..136  
 FT /tag- b  
 FT mat\_peptide 137..1510  
 FT /tag- c  
 FT U55599709-A.  
 PN 04-FEB-1997.  
 PD 28-SEP-1989; 413947.  
 PR 28-SEP-1989; US-413947.  
 PR (SALK) SALK INST BIOLOGICAL STUDIES.  
 PI Lindstrom JM, Schoepfer RD;  
 DR WPI: 97-118297/11.  
 DR P-PSDB; W12368.  
 PT New isolated neuronal alpha-bungarotoxin-binding protein DNA - used  
 PT to screen cholinergic agents and other drugs which may affect ligand  
 PT binding, ion channel or other activities of the protein.  
 PS Claim 1; Fig 2A-B; 18pp: English.  
 CC 2 cDNA clones (T59196 and T59197) respectively code for the alpha1  
 CC subunit (W12368) and alpha2 subunit (W12369) of chick neuronal  
 CC alpha-bungarotoxin binding protein (ABBP). They were isolated from  
 CC an 18-day embryo chick brain cDNA library using a probe (see also  
 CC T59198) based on the N-terminal amino acid sequence of chicken  
 CC brain ABBP. The probe isolated partial clone pch29-1, which  
 CC encoded the N-terminal portion of alpha1. A subclone, pch29-3  
 CC (ATCC 40641), was used to rescreen the library, yielding clone  
 CC pch31-1 (ATCC 40640), which encoded the entire alpha2 sequence.  
 CC A probe based on the C-terminal region of pch31-1 was used obtain  
 CC clone pch34-1 (ATCC 40639), encoding the C-terminal portion of  
 CC alpha1. The cDNA clones can be used as probes to identify further  
 CC ABBP subunits, and in the recombinant prodn. of ABBP.  
 SQ Sequence 2769 BP; 790 A; 604 C; 804 T;  
 59.5%; Score 22; DB 1; Length 2769;  
 Query Match  
 Best Local Similarity 83.3%; Pred. No. 3.1;  
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 8 CAGGCATATTCAGAGTCTCTGCTACATCG 37  
 ||||||||||||||||||||||||||||  
 DB 498 CAGGCATATTCAGAGTCTCTGCTACATCG 527  
 RESULT 5  
 ID T48237 standard; DNA; 3496 BP.  
 AC T48237;  
 DT 09-APR-1997 (first entry)  
 DE Neuronal nicotinic acetylcholine receptor alpha-4 subunit DNA.  
 KW Neuronal nicotinic acetylcholine receptor; nAChR; neurotransmitter;  
 KW ligand-gated receptor; ds.

```

OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 232..2115
FT misc_difference 3344
FT /tag- a
FT /note- b
FT /note- "base 3344 is given as n in the
FT specification"
PN M09641876-A1.
PD 27-DEC-1996.
PE 07-JUN-1996; U09775.
PR (SIBI-) SIBIA NEUROSCIENCES INC.
PI Elliott KJ, Harpold KM;
DR WPI; 97-065463/06.
DR P-PSDB; W09023.
PT Nucleic acids encoding nicotinic acetylcholine receptor sub-units -
used in screening to determine the effect of drugs on the receptor
disclosure; Page 58-61; 108pp; English.
CC A DNA sequence (T48237) codes for the alpha-4 subunit (W09023) of
the human neuronal nicotinic acetylcholine receptor (nAChR). Host
cells, esp. mammalian cells or amphibian oocytes, carrying alpha-4
nucleic acids, opt. in combination with other alpha and/or beta
subunit nucleic acids (see also T48232-36, T48238-41), express
recombinant nAChR subunits useful for identifying cpds. that
modulate the activity of human nAChRs.
SQ Sequence 3496 BP; 590 A; 1215 C; 1084 G; 606 T;

Query Match
Best Local Similarity 56.8%; Score 21; DB 1; Length 3496;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ACCGACACAGGCAATTCAGAGTCTCCTGCTACATCG 37
DB 685 ACTCCCGCGGCATTACAGAGCTCTGACAGATCG 721

RESULT 6
ID T59527
AC T59527 standard; DNA; 2082 BP.
DE 05-OCT-1997 (first entry)
KW Alpha4 subunit of normal nAChR.
KW nAChR; mutation; autosomal dominant nocturnal frontal lobe epilepsy;
KW ADNFLE; neuronal acetylcholine receptor; ss.
Key Location/Qualifiers
FT misc_rna 902..1168
FT /tag- a
FT /note- "ion channel"
FT mutation 1016..1018
FT /tag- b
FT /note- "TCC 248 TTC is the ADNFLE-
FT inducing mutation (see T59528)"
PN AU9656247-A.
PD 09-JAN-1997.
PE 28-JUN-1996; 056247.
PR 28-JUN-1996; AU-003840.
PA (UYBO-) UNIV BONN.
PA (WOME-) UNIV MELBOURNE.
PA (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
PI Berkovic SF, Mulley JC, Phillips HA,
PI Scheffer IE, Steinleink, Sutherland GR, Wallace RH,
DR WPI; 97-100506/10.
DR P-PSDB; W11824.
PT CHRNA4 gene encodes neuronal nicotinic acetylcholine receptor
PT fragment - used in diagnosis of autosomal dominant nocturnal frontal
PT lobe epilepsy
PS Claim 5; Fig 13; 20pp; English.
CC The DNA sequence of the normal nAChR is given in T59527,
CC the DNA sequence of the ADNFLE-inducing mutant is given in T59528.
CC Mutations at codon 248, which have the effect of replacing
CC serine by another amino acid (e.g. phenylalanine) in the alpha
CC subunit of nAChR have been found to be associated with ADNFLE.
CC The primers given in T59529 and T59530 were used in the
CC amplification of part of exon 5.
SQ Sequence 2082 BP; 367 A; 768 C; 592 G; 355 T;

Query Match
Best Local Similarity 56.8%; Score 21; DB 1; Length 2082;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ACCGACACAGGCAATTCAGAGTCTCCTGCTACATCG 37
DB 632 ACTCCCGCGGCATTACAGAGCTCTGACAGATCG 668

RESULT 7
ID T59528
AC T59528 standard; DNA; 2082 BP.
DE 05-OCT-1997 (first entry)
KW Alpha4 subunit of mutant nAChR (TCC 248 TTC).
KW nAChR; mutation; autosomal dominant nocturnal frontal lobe epilepsy;
KW ADNFLE; neuronal acetylcholine receptor; ss.
Key Location/Qualifiers
FT misc_rna 902..1168
FT /tag- a
FT /note- "ion channel"
FT mutation 1016..1018
FT /tag- b
FT /note- "TCC 248 TTC is the ADNFLE-
FT inducing mutation (see T59527)"
PN AU9656247-A.
PD 09-JAN-1997.
PE 28-JUN-1996; 056247.
PR 28-JUN-1996; AU-003840.
PA (UYBO-) UNIV BONN.
PA (WOME-) UNIV MELBOURNE.
PA (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
PI Berkovic SF, Mulley JC, Phillips HA,
PI Scheffer IE, Steinleink, Sutherland GR, Wallace RH,
DR WPI; 97-100506/10.
DR P-PSDB; W11825.
PT CHRNA4 gene encodes neuronal nicotinic acetylcholine receptor
PT fragment - used in diagnosis of autosomal dominant nocturnal frontal
PT lobe epilepsy
PS Claim 6; Fig 13; 20pp; English.
CC The DNA sequence of the normal nAChR is given in T59527,
CC the DNA sequence of the ADNFLE-inducing mutant is given in T59528.
CC Mutations at codon 248, which have the effect of replacing
CC serine by another amino acid (e.g. phenylalanine) in the alpha
CC subunit of nAChR have been found to be associated with ADNFLE.
CC The primers given in T59529 and T59530 were used in the
CC amplification of part of exon 5.
SQ Sequence 2082 BP; 367 A; 767 C; 592 G; 356 T;

Query Match
Best Local Similarity 56.8%; Score 21; DB 1; Length 2082;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ACCGACACAGGCAATTCAGAGTCTCCTGCTACATCG 37
DB 632 ACTCCCGCGGCATTACAGAGCTCTGACAGATCG 668

RESULT 8
ID V12196
AC V12196 standard; CDNA; 2363 BP.

```

```

CC serine by another amino acid (e.g. phenylalanine) in the sixth
CC amino acid of the transmembrane domain 2 (M2) of the alpha4
CC subunit of nAChR have been found to be associated with ADNFLE.
CC The primers given in T59529 and T59530 were used in the
CC amplification of part of exon 5.
SQ Sequence 2082 BP; 367 A; 768 C; 592 G; 355 T;

Query Match
Best Local Similarity 56.8%; Score 21; DB 1; Length 2082;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ACCGACACAGGCAATTCAGAGTCTCCTGCTACATCG 37
DB 632 ACTCCCGCGGCATTACAGAGCTCTGACAGATCG 668

RESULT 7
ID T59528
AC T59528 standard; DNA; 2082 BP.
DE 05-OCT-1997 (first entry)
KW Alpha4 subunit of mutant nAChR (TCC 248 TTC).
KW nAChR; mutation; autosomal dominant nocturnal frontal lobe epilepsy;
KW ADNFLE; neuronal acetylcholine receptor; ss.
Key Location/Qualifiers
FT misc_rna 902..1168
FT /tag- a
FT /note- "ion channel"
FT mutation 1016..1018
FT /tag- b
FT /note- "TCC 248 TTC is the ADNFLE-
FT inducing mutation (see T59527)"
PN AU9656247-A.
PD 09-JAN-1997.
PE 28-JUN-1996; 056247.
PR 28-JUN-1996; AU-003840.
PA (UYBO-) UNIV BONN.
PA (WOME-) UNIV MELBOURNE.
PA (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
PI Berkovic SF, Mulley JC, Phillips HA,
PI Scheffer IE, Steinleink, Sutherland GR, Wallace RH,
DR WPI; 97-100506/10.
DR P-PSDB; W11825.
PT CHRNA4 gene encodes neuronal nicotinic acetylcholine receptor
PT fragment - used in diagnosis of autosomal dominant nocturnal frontal
PT lobe epilepsy
PS Claim 6; Fig 13; 20pp; English.
CC The DNA sequence of the normal nAChR is given in T59527,
CC the DNA sequence of the ADNFLE-inducing mutant is given in T59528.
CC Mutations at codon 248, which have the effect of replacing
CC serine by another amino acid (e.g. phenylalanine) in the alpha
CC subunit of nAChR have been found to be associated with ADNFLE.
CC The primers given in T59529 and T59530 were used in the
CC amplification of part of exon 5.
SQ Sequence 2082 BP; 367 A; 767 C; 592 G; 356 T;

Query Match
Best Local Similarity 56.8%; Score 21; DB 1; Length 2082;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ACCGACACAGGCAATTCAGAGTCTCCTGCTACATCG 37
DB 632 ACTCCCGCGGCATTACAGAGCTCTGACAGATCG 668

RESULT 8
ID V12196
AC V12196 standard; CDNA; 2363 BP.

```

16-APR-1998.  
PF 09-OCT-1997; U18802.  
PR 10-OCT-1996; US-028191.  
PA (RGC ) UNIV CALIFORNIA.  
PI Bodeau J, Kaloshian I, Milligan S, Williamson VM,  
P1 Yaghoobi J;  
Wpi: 98-240529/21.  
DR P-FSDB: W55974, W55975.  
PT Nucleic acids encoding M1 polypeptide(s) conferring nematode  
resistance - useful to produce transgenic plants resistant to these  
and other pests, and in marker-aided selection to assess cultivars  
for resistance  
Claim 11: Page 24-39; 55pp; English.  
CC This is the nucleotide sequence of the tomato M1 locus associated  
with nematode resistance, obtained from a bacterial artificial  
chromosome (BAC). M1 was localised by genetic analysis to a region  
of the tomato genome of about 65 kb. DNA corresponding to this  
region was cloned into BAC vectors. Sequence analysis of a 52 kb  
BAC3 insert identified 3 open reading frames, one of which is  
probably a pseudogene. By RNA blot analysis, transcripts of  
approximately 4 kb corresponding to copy 1 and copy 2 were found in  
both resistant and susceptible tomato roots and in leaves of  
resistant but not susceptible plants. cDNA sequences corresponding  
to full-length transcripts of copy 1 (see V26082) and copy 2 (see  
V26083) were obtained. The encoded polypeptides (see W55974-75)  
are 91% identical and contain structural features similar to/know  
plant resistance genes (R genes) of the nucleotide binding site/  
leucine-rich repeat (NBS/LRR) family. A recombinant expression  
cassette comprising an M1 polynucleotide and an operably linked  
plant promoter can be used to enhance nematode resistance in plants  
especially tomatoes (claimed). Transgenic plants can also be  
constructed using a M1 promoter with heterologous genes: the M1  
promoters can be used to express a variety of genes in the same  
temporal and spatial patterns and at similar levels to resistance  
genes.  
SQ Sequence 51952 BP; 17728 A; 8444 C; 8647 G; 17111 T;

Query Match 56.2%; Score 20.8; DB 1; Length 51952;  
Best Local Similarity 78.1%; Pred. No. 16;  
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps

QY 5 ACACAGCATATTCAGAGTCTCCTGCTACATC 36  
||| ||||| ||||| ||||| |||||  
Db 30281 ACAAGCGCATTCACAGATTCTGTACTGACAGC 30250

RESULT 10  
Q90387 standard; CDNA; 2277 BP.

ID Q90387;  
AC 090387;  
DT 30-NOV-1995 (first entry)  
DE Alpha 2 subunit of a human neuronal nicotinic acetylcholine receptor;  
KW Human nAChR; neuronal nicotinic acetylcholine receptor;  
KM neurotransmitter; ss.  
OS Homo sapiens.  
FH key location/Qualifiers  
FT cds 166..1755  
FT /\*tag= a

W09513299-A:  
PN 18-MAY-1995.  
PD 08-NOV-1994; U12859.  
PE 08-NOV-1993; US-149503.  
PR (SALK ) INST BIOTECHNOLOGY IND ASSOC.  
PA Eliott KJ, Ellis SB, Harpold MM;  
PI Wpi: 95-194036/75.  
DR P-FSDB: R73966.  
PT New human neuronal nicotinic acetylcholine receptor alpha2 subunit DNA  
used to develop prods. for detection, diagnosis and therapy and for  
modulating activity  
Claim 2; Page 43-46; 54pp; English.  
PS DNA encoding the human nAChR alpha2 subunit was isolated from a  
human thalamus tissue cDNA library using corresp. rat cDNA. The

CC Insert of one clone cdt. was ligated with the insert of another  
CC clone to generate a full-length alpha 2 subunit cDNA. The DNA can be  
CC used to identify function NACHRs. Cells contg. the DNA can be  
CC for screening to identify cpds. which modulate the activity of human  
CC nACHRs. The human nACHR alpha 2 subunit can be used to product  
CC antibodies which can be used in immunohistochemistry, diagnosis and  
CC therapy. The nucleic acids can be used for analysing disease states  
CC and creating animal models.  
SO Sequence 2277 BP; 451 A; 696 C; 643 G; 487 T;

Query Match 53.0%; Score 19.6; DB 1; Length 2277;  
Best Local Similarity 73.5%; Pred. No. 29;  
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 CACACAGCATATTCAGAGTCTCTGCTACATCG 37  
nb 688 CCCCCGCCATCTACAGAGCTCTCTGACATCG 721

RESULT 11

ID T48235 standard; DNA: 2664 BP.  
AC T48235;  
DT 09-APR-1997 (first entry)  
DE Neuronal nicotinic acetylcholine receptor alpha-2 subunit DNA.  
KW Neuronal nicotinic acetylcholine receptor; nACHR; neurotransmitter;  
OS Homo sapiens.  
FH Key  
FT Location/Qualifiers  
FT cds  
FT 555..2141  
FT /\*tag= a

DR W09641876-A1.  
DR 27-DEC-1996.  
PR 07-JUN-1996; U09775.  
PR (SIBI-) SIBIA NEUROSCIENCES INC.  
PI Elliott KJ, Harpold MW;  
DR WPI; 97-065463/06.  
P-PSDB: W09021.  
PT Nucleic acids encoding nicotinic acetylcholine receptor sub-units -  
PT used in screening to determine the effect of drugs on the receptor  
CC A DNA sequence (T48235) codes for the alpha-2 subunit (W09021) of  
CC the human neuronal nicotinic acetylcholine receptor (nACHR). Host  
CC cells, esp. mammalian cells or amphibian oocytes, carrying alpha-2  
CC subunit nucleic acids (see also T48232-34, T48236-41), express  
CC recombinant nACHR subunits useful for identifying cpds. that  
CC modulate the activity of human nACHRs.  
SO Sequence 2664 BP; 518 A; 815 C; 743 G; 588 T;

Query Match 53.0%; Score 19.6; DB 1; Length 2664;  
Best Local Similarity 73.5%; Pred. No. 30;  
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 CACACAGCATATTCAGAGTCTCTGCTACATCG 37  
Db 1077 CCCCCGCCATCTACAGAGCTCTCTGACATCG 1110

RESULT 12

ID V12199 standard; cDNA: 2277 BP.  
AC V12199;  
DT 14-MAY-1998 (first entry)  
DE Human neuronal nicotinic acetylcholine receptor alpha-2 subunit cDNA.  
KW Human; neuronal nicotinic acetylcholine receptor; alpha-2 subunit;  
OS Homo sapiens.  
FH Key  
FT Location/Qualifiers  
FT cds  
FT 106..1175

FT /\*tag= a  
FT /product= "neuronal nicotinic acetylcholine receptor  
FT alpha-2 subunit"

FN W09420617-A2.  
PD 15-SEP-1994.  
PR 08-MAR-1994; U02447.  
PR 08-MAR-1993; US-028031.  
PA (SALK-) SALK INST BIOTECHNOLOGY IND ASSOC.  
PI (SIBI-) SIBIA NEUROSCIENCES INC.  
PI Elliott KJ, Ellis SB, Harpold MW;  
DR WPI; 94-303024/37.  
P-PSDB: W44155.

PT Human neuronal nicotinic acetylcholine receptor subunits and DNA -  
PT also transformed cells useful for screening cpds. which modulate  
PT activity of the receptor  
PS Disclosure; Page 67-68; 99pp; English.  
CC The present sequence encodes a human neuronal nicotinic acetylcholine  
CC receptor (NACHR) subunit. The cells expressing the alpha and/or beta  
CC NACHR subunits may be used in a method of screening compounds to  
CC identify any which modulate the activity of human neuronal NACHR.  
CC and expression density of various subunits in normal vs diseased brain  
CC tissues. Testing of single receptor subunits or specific receptor  
CC subunit combinations with a variety of potential agonists or antagonists  
CC provides information with respect to the function and activity of the  
CC individual subunits and should lead to the identification and design of  
CC compounds that are capable of very specific interaction with one or  
CC more receptor subtypes. The resulting drugs should exhibit fewer  
CC unwanted side effects than drugs identified e.g. screening with cells  
CC that express a variety of subtypes.  
SO Sequence 2277 BP; 451 A; 695 C; 643 G; 486 T;

Query Match 53.0%; Score 19.6; DB 1; Length 2277;  
Best Local Similarity 73.5%; Pred. No. 29;  
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 CACACAGCATATTCAGAGTCTCTGCTACATCG 37  
Db 688 CCCCCGCCATCTACAGAGCTCTCTGACATCG 721

RESULT 13

ID V28636 standard; cDNA: 377 BP.  
AC V28636;  
DT 29-JUL-1998 (first entry)  
DE Human aspartic protease partial cDNA EST 857644 SRO ID NO:14.  
KW Human; aspartic protease; EST; endothelin; serum amyloid A protein;  
KW pro-opiomelanocortin prohormone; ss.  
OS Homo sapiens.  
PN W09811236-A1.  
PD 19-MAR-1998.  
PR 09-SEP-1997; G02426.  
PR 11-SEP-1996; GB-018966.  
PA (SMIK-) SMITHKLINE BEECHAM PLC.  
PI (UYMA-) UNIV MALES.  
PI Hill J, Kay J, Powell D;  
DR WPI; 98-207396/18.

PT Human aspartic protease and related DNA - which may be involved in  
PT processing of endothelin and pro-opiomelanocortin pro-hormone(s)  
PS Disclosure; Page 23; 32pp; English.  
CC The present sequence represents a partial cDNA sequence from a human  
CC aspartic protease, isolated from a cDNA library of human origin. The  
CC from human cDNA libraries having partial DNA sequences which encode  
CC aspartic protease. Compounds which inhibit aspartic protease, especially  
CC used in therapy where needed. Important functions of aspartic proteases  
CC are the processing of endothelin and pro-opiomelanocortin prohormones.  
CC They may also be involved in the processing of serum amyloid A protein.  
SO Sequence 377 BP; 62 A; 97 C; 121 G; 89 T;



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:26 ; Search time 2825.05 Seconds

(without alignments)  
25.834 Million cell updates/sec

Title: US-08-956-518a-100

Perfect score: 37  
1 ACCCACACAGGCATATTCAGAGTCTGCTACATCG 37

Scoring table: IDENTITY\_NUC

arched: 2546578 seqs, 986266752 residues

Database:

EST:  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	37	100.0	261	21	F08364	F08364 HSCZS111 n
2	37	100.0	502	39	AA889613	AA889613 AK28611.s
3	37	100.0	497	43	AI241498	AI241498 qh69c08.x
4	37	100.0	476	45	AI368029	AI368029 q945a03.x
5	36	97.3	421	25	N71634	N71634 za1c07.s1
6	29	78.4	405	26	W93369	W93369 zh46a03.s1
7	27	73.0	106	26	W86805	W86805 zh64g04.s1
8	26.2	70.8	560	26	W52861	W52861 zc03e02.r1
9	22.8	61.6	293	23	R88973	R88973 ym97b10.r1
10	22.4	60.5	511	47	AI509992	AI509992 m119h09.y
11	22.2	60.0	426	22	H01869	H01869 y132a12.r1
12	22.2	60.0	411	22	R32633	R32633 yh73h02.r1
13	22.2	60.0	393	22	R71355	R71355 y154f12.r1
14	22	59.5	418	36	AA621981	AA621981 nq24c12.s
15	21.8	58.9	395	46	AA664815	AA664815 U1-R-CO-g
16	21.2	57.3	416	22	R82535	R82535 y119b06.r1
17	21.2	57.3	409	25	W03952	W03952 za62c08.r1
18	20.8	56.2	1161	26	W41959	W41959 mce6b10.r1
19	20.8	56.2	728	50	AA050499	AA050499 m320f02.r
20	20.6	55.7	337	20	AI683454	AI683454 tw51f01.x
21	20.6	55.7	337	20	AA82609	AA82609 c634d07.s
22	20.6	55.7	341	39	AA82609	AA82609 c634d07.s
23	20.2	54.6	568	35	AA537135	AA537135 vk45c09.r
24	20.2	54.6	398	40	AA992919	AA992919 o04d04.s
25	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
26	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
27	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
28	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
29	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
30	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
31	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
32	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
33	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
34	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
35	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
36	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
37	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
38	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
39	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
40	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
41	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
42	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
43	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
44	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
45	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
46	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
47	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
48	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
49	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
50	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
51	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
52	20.2	54.6	373	40	AA992919	AA992919 o04d04.s
53	20.2	54.6	373	40	AA992919	AA992919 o04d04.s

#### ALIGNMENTS

RESULT 1  
F08364/c  
LOCUS F08364 261 bp mRNA  
DEFINITION HSCZS111 normalized infant brain cDNA Homo sapiens  
ACCESSION F08364  
NID 9676930  
VERSION F08364.1 GI:676930

EST 22-FEB-1995  
c-zsa11, mRNA sequence.

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 261)  
AUTHORS Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes, M.D., Duprat, S., Houligatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pletu, G., Pouliot, Y., Sebastiani, Kaakrichis, C. and Tessier, A.  
IMAGE: molecular integration of the analysis of the human genome and its expression  
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
95277534

JOURNAL MEDLINE  
COMMENT

CONTACT: Genethon  
Genexpress-Genethon  
Genethon Centre de recherche sur le Genome Humain  
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE  
Tel: 33169472800  
Fax: 33160778698  
Email: genexpress@genethon.fr  
Single read.  
Genexpress-library-Idt: C; Genexpress-sequence-Idt: y1c-zsall  
Insert length: 664 Std Error: 0.00  
Seq primer: (-21)M13 universal  
High quality sequence stop: 436.  
Location/Qualifiers

FEATURES  
source  
1..261  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="c-zsall"  
/clone-lib="normalized infant brain cDNA"  
/sex="female"  
/tissue="total brain"  
/dev\_stage="3 months old"  
/note="Organ: brain; Vector: lafmid BA; Site: 1: HindIII;  
Site 2: NotI; sex: female; dev\_stage=3 months old;  
Isolate-muscular atrophy patient; tissue-type-total  
brain; total mRNA was oligo-(dt) primed and directionally  
cloned 5' -> 3' into the HindIII -> NotI sites of the  
laifmid BA vector. Clone library from B. Soares, Psychiatry  
Dept. Columbia University, USA. Normalization\_method:  
Bento Soares, P.N.A.S in press

BASE COUNT  
71 a 65 c 60 g 63 t 2 others

ORIGIN  
Query Match 100.0%; Score 37; DB 21; Length 261;  
Best Local Similarity 100.0%; Pred. No. 1.2e-05;  
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAGCATATTCAGAGTCCGCTACATCG 37  
|||||  
Db 253 ACCCACACAGCATATTCAGAGTCCGCTACATCG 217

RESULT 2  
AA889613 502 bp mRNA EST 04-JAN-1999  
LOCUS AA889613  
DEFINITION ak28611.s1 Soares, testis\_NHT Homo sapiens cDNA clone IMAGE:1407308  
3' similar to gb:170297 NEURONAL ACETYLCHOLINE RECEPTOR PROTEIN,  
ALPHA-7 CHAIN (HUMAN);, mRNA sequence.  
ACCESSION AA889613  
NID 93016492  
VERSION AA889613.1 GI:3016492  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 502)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
COMMENT

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 9, 1998 this sequence version replaced gi:936210.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www.bio.lnl.gov/Db/rp/image/image.html

FEATURES  
source  
Insert length: 376 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 255.  
Location/Qualifiers

1..502  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1407308"  
/clone-lib="Soares-testis\_NHT"  
/sex="male"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from Clontech  
Laboratories, Inc., and primed with a Not I - oligo(dT)  
primer [5']  
TGTACCATCTGAGTGGAGCGCCGCCCATTTTCTTTT 3']  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
was through one round of normalization to cots, and was  
constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT  
128 a 98 c 122 g 154 t

ORIGIN  
Query Match 100.0%; Score 37; DB 39; Length 502;  
Best Local Similarity 100.0%; Pred. No. 1.3e-05;  
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAGCATATTCAGAGTCCGCTACATCG 37  
|||||  
Db 319 ACCCACACAGCATATTCAGAGTCCGCTACATCG 355

RESULT 3  
A1241498 497 bp mRNA EST 04-NOV-1998  
LOCUS A1241498  
DEFINITION q169d08.x1 Soares, fetal\_liver\_spleen\_INFIS.S1 Homo sapiens cDNA  
clone IMAGE:1849935 3' similar to gb:170297 NEURONAL ACETYLCHOLINE  
RECEPTOR PROTEIN, ALPHA-7 CHAIN (HUMAN);, mRNA sequence.  
ACCESSION A1241498  
NID 93636895  
VERSION A1241498.1 GI:3836895  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 497)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Aug 21, 1998 this sequence version replaced.  
Contact: Robert Strausberg, Ph.D.





```

1..405
/organism="Homo sapiens"
/db_xref="GDB:1123554"
/db_xref="taxon:9606"
/map="19p12-p13.1"
/clone="IMAGE:415084"
/clone_1fb="Soares.fetal_liver.spleen.lNPLS_S1"
/sex="male"
/dev_stage="70 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pTR10d (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI"
This is a subtracted version of the original Soares fetal
liver spleen lNPLS library. 1st strand cDNA was primed
with a Pac I - oligo(dt) primer [5'
AATCGAGAAATTATTAATGAACCTTTTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pTR13 vector. Library

```

## FEATURES

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel.: 314 286 1800  
Fax: 314 286 1810  
Email: [estwatson.wustl.edu](mailto:estwatson.wustl.edu)  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
Trace considered overall poor quality  
Insert length: 1097 Std Error: 0.00  
Seq primer: mob.RGCA+ET  
High quality sequence stop: 1.  
Insertion/Deletion

```

1..106
/organism="Homo sapiens"
/db_xref="GDB:135356"
/db_xref="taxon:9606"
/clone IMAGE:418886"
/clone_lib="Scars fetal_liver_spleen_INFLS_S1"
/sex="male"
/dev_stage="20 week post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/vector="p7T3D (Pharmacia)
note=Organ: Liver and Spleen; Vector: p7T3D (Pharmacia)
with a modified polylinker. Site_1: Pac I; Site_2: Eco RI
This is a subtracted version of the original Scars fetal
liver spleen INFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dC) primer [5'
ACCTGCAGCAATTATTAATGAAGCTCTTTTTTTTTTTT 3'],
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified p7T3 vector. Library

```

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
Insert Size: 2357  
High quality sequence stops: 241 Source: IMAGE Consortium, LINT  
This clone is available royalty-free through LINT; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 2357 Std Error: 0.00  
Seq primer: M3RPl  
High quality sequence stop: 241.  
Location/Qualifiers  
1.293  
/organism="Homo sapiens"  
/db\_xref="GDB:587538"  
/db\_xref="taxon:9606"  
/clone="IMAGE:166843"  
/clone\_lib="Soares adult brain N2b4HB5Y"  
/sex="Male"  
/dev\_stage="55-year old"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: brain; Vector: pRT73D (Pharmacia) with a  
modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TCTTACATCTCGAATGCGAGCGCGCCGCTTTTCTTTTCTTTT 3'], RI  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pRT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 53. Library constructed by Bento  
Soares and M.Felipe Donaldo. The adult brain RNA was  
provided by Dr. Donald H. Gilden. Tissue was acquired  
17-18 hours after death which occurred in consequence of a

ruptured aortic aneurysm. RNA was prepared from a pool of tissues representing the following areas of the brain: frontal, parietal, temporal and occipital cortex from the left and right hemispheres, subcortical white matter, basal ganglia, thalamus, cerebellum, midbrain, pons and medulla."

BASE COUNT 49 a 90 c 80 g 69 t 5 others  
ORIGIN

Query Match 61.6%; Score 22.8; DB 23; Length 293;  
Best Local Similarity 77.1%; Pred. No. 8.1;  
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 2 CCCACAGGCGATATTCAGAGTCTGCTGATAC 36  
Db 118 CCCACAGGCGATATTCAGAGTCTGCTGATC 152

---SULT 10  
09992

DEFINITION A1509992 511 bp mRNA EST 12-MAR-1999  
m19h09.y1 Soares mouse embryo NDM3.5 14.5 Mus musculus cDNA  
clone IMAGE:476609 5' similar to gb:U15647\_cds1 Mus musculus  
(MUSE);, mRNA sequence.

ACCESSION A1509992  
NID 94408897

VERSION A1509992.1 GI:4408897  
KEYWORDS EST.

ORGANISM house mouse.  
SOURCE

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 511)

AUTHORS Matra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,  
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,  
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
Waterston,R. and Wilson,R.

REFERENCE The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)

JOURNAL On Jun 5, 1998 this sequence version replaced gi:3188684.

COMMENT

Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
This read is a RESEQUENCE of a previously sequenced mouse clone  
This read has been verified (found to hit its original self in the  
correct orientation)  
Putative full length read  
vector to vector length is 577  
Seq primer: -40RP from Glibco  
High quality sequence stop: 447.

Location/Qualifiers  
1..511  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_image="IMAGE:476609"  
/clone\_lib="Soares mouse embryo NDM3.5 14.5"  
/sex="unknown"  
/tissue\_type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"  
/lab\_host="DH10B"  
/note="Vector: pRT73P-Pac (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo (dT) primer [5'  
TGTACCAATCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTT

FEATURES  
source

BASE COUNT 186 a 135 c 76 g 113 t 1 others  
ORIGIN

Query Match 60.5%; Score 22.4; DB 47; Length 511;  
Best Local Similarity 81.2%; Pred. No. 12;  
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 3 CCACAGGCGATATTCAGAGTCTGCTGATAC 34  
Db 443 CCACAGGCGATATTCAGAGTCTGCTGATAC 474

RESULT 11  
H01869

DEFINITION H01869 426 bp mRNA EST 19-JUN-1995  
Y132a12.r1 Soares placenta Nb2HP Homo sapiens cDNA clone  
IMAGE:150430 5', mRNA sequence.

ACCESSION H01869  
NID 9864802

VERSION H01869.1 GI:864802  
KEYWORDS EST.

ORGANISM human.  
SOURCE

REFERENCE Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 426)

AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hulman,M., Kucaba,T., Le,M., Lennon,G., Matra,M.,  
Parsons,J., Rikkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Treviskis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Wilson,R.

REFERENCE The WashU-Merck EST Project  
Unpublished (1995)

JOURNAL On May 9, 1995 this sequence version replaced gi:804305.

COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: est@wustl.edu  
Insert Size: 1235  
High quality sequence stops: 322  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert length: 1235 Std Error: 0.00  
Seq primer: M13RP1  
High quality sequence stop: 322.

Location/Qualifiers  
1..426  
/organism="Homo sapiens"  
/db\_xref="GDB:562415"  
/db\_xref="taxon:9606"  
/clone\_image="IMAGE:150430"  
/clone\_lib="Soares placenta Nb2HP"  
/sex="Female"  
/dev\_stage="placenta obtained at birth (full term)"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: placenta. Vector: pRT73D (Pharmacia) with a  
modified polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo (dT) primer [5'  
AAGTGGAAGATTCGCGCGCGGAGGAAATTTTTTTTTTTTTTTTTT 3']  
double-stranded cDNA was ligated to Eco RI adaptors

FEATURES  
source

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fátima Bernaldo. "

BASE COUNT 69 a 136 c 122 g 95 t 4 others

ORIGIN

Query Match 60.0%; Score 22.2; DB 22; Length 426;  
Best Local Similarity 77.1%; Pred. No. 15;  
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 CCCACAGAGCATATTCAGAGTTCCTGCTACATC 36  
Db 33 CCCACAGAGCATATTCAGAGTTCCTGCTACATC 67

RESULT 12  
3633

DEFINITION R32633 411 bp mRNA EST 28-APR-1995  
Y173H02.r1 Soares Placenta NB2HP Homo sapiens cDNA clone  
IMAGE:135411 5', mRNA sequence.

ACCESSION R32633  
NID 9788476  
VERSION R32633.1 GI:788476  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
AUTHORS Hallier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Wilson,R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1281  
High quality sequence stops: 244  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 1281 Std Error: 0.00  
Seq primer: M13RP1  
High quality sequence stop: 244.  
Location/Qualifiers

FEATURES  
source

1..411  
/organism="Homo sapiens"  
/db\_xref="GDB:541344"  
/db\_xref="taxon:9606"  
/clone="IMAGE:135411"  
/clone.lib="Soares placenta NB2HP"  
/sex="Female"  
/dev\_stage="Placenta obtained at birth (full term)"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
AAGTCGAAGAATTCGCGCGCAGAGATTTTCTTTTCTTTT 3']  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fátima Bernaldo. "

BASE COUNT 74 a 123 c 117 g 94 t 3 others

ORIGIN

Query Match 60.0%; Score 22.2; DB 22; Length 411;  
Best Local Similarity 77.1%; Pred. No. 15;  
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 CCCACAGAGCATATTCAGAGTTCCTGCTACATC 36  
Db 33 CCCACAGAGCATATTCAGAGTTCCTGCTACATC 67

RESULT 13  
R71355

LOCUS R71355 393 bp mRNA EST 01-JUN-1995  
DEFINITION Y154F12.r1 Soares placenta NB2HP Homo sapiens cDNA clone  
IMAGE:143087 5', mRNA sequence.

ACCESSION R71355  
NID 9844872  
VERSION R71355.1 GI:844872  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
AUTHORS Hallier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Wilson,R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT On May 18, 1995 this sequence version replaced gi:811226.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1209  
High quality sequence stops: 347  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 1209 Std Error: 0.00  
Seq primer: M13RP1  
High quality sequence stop: 347.  
Location/Qualifiers

FEATURES  
source

1..393  
/organism="Homo sapiens"  
/db\_xref="GDB:552210"  
/db\_xref="taxon:9606"  
/map="17"  
/clone="IMAGE:143087"  
/clone.lib="Soares placenta NB2HP"  
/sex="Female"  
/dev\_stage="Placenta obtained at birth (full term)"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
AAGTCGAAGAATTCGCGCGCAGAGATTTTCTTTTCTTTT 3']  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fátima Bernaldo. "

BASE COUNT 68 a 129 c 113 g 79 t 4 others

Query Match 60.0%; Score 22.2; DB 22; Length 393;  
 Best Local Similarity 77.1%; Pred. No. 15;  
 Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 2 CCCACACAGGATATTCAGAGTCTCTACATC 36  
 32 CCCACACCCCATATTCATGAGTACCTGATC 66

RESULT 14  
 LOCUS AA621981/c  
 DEFINITION hg24c12.s1 NCI-CGAP Co10 Homo sapiens CDNA clone IMAGE:1144822.3'  
 similar to TR:G1339923 G1339923 PHENOL SULFOTRANSFERASE SUBUNIT ;  
 mRNA sequence.

ACCESSION AA621981  
 NID 92525857  
 VERSION AA621981.1 GI:2525857  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 418)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL On Sep 12, 1996 this sequence version replaced gi:1407520.  
 COMMENT

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,  
 M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bdip/image/image.html

Trace considered overall poor quality  
 Insert length: 1697 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 1.

FEATURES  
 Source

1. 418  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="1144822"  
 /clone\_id="NCI-CGAP\_Co10"  
 /tissue\_type="Colon tumor RER+"  
 /lab\_host="DH10B"  
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a  
 modified polylinker; 1st strand cDNA was prepared from  
 RER colon tumor; and was then primed with a Not I -  
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco  
 RI adaptors (Pharmacia), digested with Not I and cloned  
 into the Not I and Eco RI sites of the modified pT73  
 vector. Library is normalized. Library was constructed by  
 Bento Soares and M. Fatima Bonaldo (N-Soares4)."

BASE COUNT 107 a 82 c 116 g 113 t

ORIGIN

Query Match 59.5%; Score 22; DB 36; Length 418;  
 Best Local Similarity 83.3%; Pred. No. 18;  
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Db 1 ACCACACAGGATATTCAGAGTCTCTGC 30  
 99 ACCGACACGATTCAGATTCAGAGATCCCTGC 70

RESULT 15  
 LOCUS AA964815/c  
 DEFINITION UI-R-C0-gx-a-10-0-UI s1 UI-R-C0 Rattus norvegicus CDNA clone  
 UI-R-C0-gx-a-10-0-UI 3', mRNA sequence.

ACCESSION AA964815  
 NID 94279689  
 VERSION AA964815.1 GI:4279689  
 KEYWORDS EST.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 395)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 Bonaldo, M.F., Lennon, G. and Soares, M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 9704447

JOURNAL MEDLINE  
 COMMENT On May 18, 1998 this sequence version replaced gi:3138307.

Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565

Email: mscares@blue.weeg.uiowa.edu  
 The sequence tag present in the cDNA between the NotI site and the  
 oligo-dT track served to identify it as a clone from the normalized  
 adult kidney library. cDNA library preparation: M. Fatima Bonaldo,  
 Ph.D. Clone distribution: clones will be available through Research  
 Genetics The following repetitive elements were found in this cDNA  
 sequence: 1-54, >R1 richlow\_complexity 57-143,  
 >(GAA)n#simple\_repeat  
 Seq primer: M13 forward.

FEATURES  
 Source

1. 395  
 /organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /map\_6p21.3: 5q33  
 /clone\_id="UI-R-C0-gx-a-10-0-UI"  
 /clone\_lib="UI-R-C0"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C0  
 library is a subtracted library derived from the UI-R-A1  
 and UI-R-E1 libraries. The UI-R-A1 library consisted of a  
 mixture of individually tagged normalized libraries  
 constructed from rat placenta, adult lung, brain, liver,  
 kidney, heart, spleen, ovary, and muscle. The UI-R-E1  
 library consisted of a mixture of individually tagged  
 normalized libraries constructed from 8, 12 and 18-day  
 embryo. The tag is a string of 3-5 nucleotides present  
 between the Not I site and the oligo-dT track which  
 allows identification of the library of origin of a clone  
 within the mixture. The subtracted library (UI-R-C0) was  
 constructed as follows: PCR amplified cDNA inserts from a  
 pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had  
 been derived was used as a driver in a hybridization with  
 the pooled UI-R-A1 and UI-R-E1 library in the form of  
 single-stranded circles. The remaining single-stranded  
 circles (subtracted library) was purified by  
 hydroxyapatite column chromatography, converted to  
 double-stranded circles and electroporated into DH10B  
 bacteria (Life Technologies) to generate the UI-R-C0  
 library. This procedure has been previously described  
 (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,  
 1996)."

BASE COUNT 71 a 76 c 108 g 140 t  
ORIGIN

Query Match 58.9%; Score 21.8; DB 46; Length 395;  
Best Local Similarity 78.8; Pred. No. 21;  
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CACACAGGCATATTCAGAGTTCCTGCTACATC 36  
||||| | | ||||| ||||| |||  
Db 223 CACACAGCCAGAGACAAGATCCTGCCACAGC 191

Search completed: September 17, 1999, 21:28:29  
Job time: 14293 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:04:23 ; Search time 1962.17 Seconds  
(without alignments)  
635.358 Million cell updates/sec

Title: US-08-956-518a-101

Perfect score: 392

Sequence: 1 AGAAGCAAGAGAGAGTAG.....CTGACGCTCCGGACTCAAC 392

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database:

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_pl1: \*  
8: gb\_pl2: \*  
9: gb\_pl3: \*  
10: gb\_pr2: \*  
11: gb\_pr3: \*  
12: gb\_to: \*  
13: gb\_st: \*  
14: gb\_sts: \*  
15: gb\_sy: \*  
16: gb\_un: \*  
17: gb\_vl: \*  
18: em\_fun: \*  
19: em\_htg: \*  
20: em\_hum1: \*  
21: em\_hum2: \*  
22: em\_in: \*  
23: em\_om: \*  
24: em\_or: \*  
25: em\_ov: \*  
26: em\_pat: \*  
27: em\_ph: \*  
28: em\_pl: \*  
29: em\_ro: \*  
30: em\_sts: \*  
31: em\_sy: \*  
32: em\_un: \*  
33: em\_vl: \*  
34: gb\_htg2: \*  
35: gb\_htg2: \*  
36: gb\_in1: \*  
37: gb\_in2: \*  
38: em\_ba1: \*  
39: em\_ba2: \*  
40: em\_hum3: \*  
41: em\_hum4: \*  
42: gb\_pr4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	392	100.0	392	11	AF029837	Homo sapi

2	72	18.4	1876	10	HSB2436	U62436 Human nicot
3	71.8	18.3	168819	35	AC007686	AC007686 Homo sapi
4	70	17.9	342379	34	CEY3986	295399 Caenorhabdi
5	69.8	17.8	2087	10	HSARA7A	X70297 H. sapiens m
6	68	17.3	1072	17	HS2IEA	M29384 Herpes slimp
7	65.8	16.8	1560	17	HS2IE	M29384 Herpes slimp
8	63.2	16.1	43058	10	HSBG1	284721 Human DNA s
9	63.2	16.1	2885	10	HOMBA3	J00184 Human alpha
10	63	16.1	154746	17	HSV2HG52	286099 Herpes slimp
11	63	15.7	154746	17	HSV2HG52	286099 Herpes slimp
12	61.6	15.7	41936	10	CH19R30879	AD000684 Homo sapi
13	61.6	15.7	14440	10	HSUSP2	Y07661 H. sapiens U
14	61.4	15.7	38532	1	SCB94	AL049628 streptomy
15	61	15.6	1927	3	CFGPAPPAM	Z12168 C. familiar
16	61	15.6	557	4	XELRGE12	K01371 X. laevis oo
17	61	15.6	3924	4	XELRGE12	J00999 X. laevis ex
18	61	15.6	7634	4	XL28SR	X59734 X. laevis 28
19	61	15.6	8153	4	XLRN01	X02995 Xenopus lae
20	61	15.6	245287	35	AC007225	AC007225 Homo sapi
21	59	15.1	3252	10	HSTAF113	Y11354 H. sapiens m
22	59	15.1	634	14	G40486	G40486 26971 Zebra
23	58.8	15.0	210636	11	AC006443	AC006443 Homo sapi
24	57.8	14.7	179436	35	AC000066	AC000066 Mus muscu
25	57.6	14.7	181884	34	HS35C18	AL022327 Homo sapi
26	57	14.5	267925	35	AC007298	AC007298 Homo sapi
27	56.8	14.5	204263	35	AC007224	AC007224 Homo sapi
28	56.4	14.4	152261	17	HE1CG	X14112 Herpes slimp
29	56.4	14.4	152261	17	HE1CG	X14112 Herpes slimp
30	56.4	14.4	6633	17	HEHSY1G3	X06461 Herpes slimp
31	56.4	14.4	26243	17	HS1US	L00036 HSV1 (strat
32	56.4	14.4	26245	17	HS1US	L00036 HSV1 (strat
33	56.2	14.3	24800	1	SC9B1	AL049727 Streptomy
34	56.2	14.3	1766	17	HS1LS17	R01835 HSV-1 (stra
35	55.8	14.2	37321	11	AC005498	AC005498 Homo sapi
36	55.6	14.2	168588	35	AC007613	AC007613 Homo sapi
37	55.6	14.2	247297	35	AC007616	AC007616 Homo sapi
38	55.2	14.1	4352	12	MMUSF2S1	U12282 Mus musculu
39	54.8	14.0	10144	10	HUMHG1Y	L17131 Human high
40	54.6	13.9	27874	12	MMU71085	U17085 Mus musculu
41	54.6	13.9	110511	14	HS180E22	AL04611 Homo sapi
42	54.2	13.8	7861	17	EVCCGA	X74312 Encephalomy
43	53.8	13.7	9701	9	AB0173353	AB017335 Homo sapi
44	53.8	13.7	42965	34	HS367G8	297634 Homo sapien
45	53.8	13.7	240229	35	AC003059	AC003059 Mus muscu

## ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	NID	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS																																			
1	AF029837	392 bp DNA	AF029837	93757792	AF029837.1	GI:3757792	human.	Homo sapiens	16-DEC-1998	U62436 Human nicot	AC007686 Homo sapi	295399 Caenorhabdi	X70297 H. sapiens m	M29384 Herpes slimp	284721 Human DNA s	J00184 Human alpha	286099 Herpes slimp	AD000684 Homo sapi	Y07661 H. sapiens U	AL049628 streptomy	Z12168 C. familiar	K01371 X. laevis oo	J00999 X. laevis ex	X59734 X. laevis 28	X02995 Xenopus lae	AC007225 Homo sapi	Y11354 H. sapiens m	G40486 26971 Zebra	AC006443 Homo sapi	AC000066 Mus muscu	AL022327 Homo sapi	AC007298 Homo sapi	AC007224 Homo sapi	X14112 Herpes slimp	X06461 Herpes slimp	L00036 HSV1 (strat	AL049727 Streptomy	R01835 HSV-1 (stra	AC005498 Homo sapi	AC007613 Homo sapi	AC007616 Homo sapi	U12282 Mus musculu	L17131 Human high	U17085 Mus musculu	AL04611 Homo sapi	X74312 Encephalomy	AB017335 Homo sapi	297634 Homo sapien	AC003059 Mus muscu

TITLE	Direct Submission					
JOURNAL	Submitted (15-OCT-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA					
FEATURES	Location/Qualifiers					
source	1. .392 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="15" /map="15q14; between D1S1031 and D1S1360" /clone="CHRNA7-p"					
Promoter	1. .392					
BASE COUNT	51 a	122 c	180 g	39 t		
ORIGIN						
Query Match	100.0%	Score 392;	DB 11;	Length 392;		
Best Local Similarity	100.0%;	Pred. No. 2,6e+46;			Gaps	0.
'atches 392;	Conservative	0;	Mismatches	0;	Indels	0;
D	1	AGAACCCAGGAGAGGTGAGACTTGCGCTTGGGACGCCCTGTGCCTGAGCAGAGGCGG	60			
	1	AGAACCCAGGAGAGGTGAGACTTGCGCTTGGGACGCCCTGTGCCTGAGCAGAGGCGG	60			
QY	61	AGGCCAGAGACCCTCGCTGCTGAGACTGGGGGTGAGAGTCCCGAGACGTACCAAGCCG	120			
Db	61	AGGCCAGAGACCCTCGCTGCTGAGACTGGGGGTGAGAGTCCCGAGACGTACCAAGCCG	120			
QY	121	GGGAGTACTCCCGCTCACACCTCGGGCTGCAAGTTCCCTTGGGTGGCCGCGAGAGCGTGG	180			
Db	121	GGGAGTACTCCCGCTCACACCTCGGGCTGCAAGTTCCCTTGGGTGGCCGCGAGAGCGTGG	180			
QY	181	CCCGGCTGAGAGATGGGGGGGGGGGAGCGGGGGCGGGGGCGGGGGCGGAGCTGTCACGTGGAG	240			
Db	181	CCCGGCTGAGAGATGGGGGGGGGGGAGCGGGGGCGGGGGCGGGGGCGGAGCTGTCACGTGGAG	240			
QY	241	AGGGCGCGGGGGGGGGGGGGGGGGGGCGCGCGCCCGGCTCTTAAAGCGCGCGAGCC	300			
Db	241	AGGGCGCGGGGGGGGGGGGGGGGGGGCGCGCGCCCGGCTCTTAAAGCGCGCGAGCC	300			
QY	301	GAGCGCGAGAGTGCCTCTGTGGCCCGCAGCGCAGAGCCCCGGSGCAGACAGAGAGSTGSAG	360			
Db	301	GAGCGCGAGAGTGCCTCTGTGGCCCGCAGCGCAGAGCCCCGGSGCAGACAGAGAGSTGSAG	360			
QY	361	C GGCGCGGCTGCTGTCAGACTCCGGGACTCAAC	392			
D	361	C GGCGCGGCTGCTGTCAGACTCCGGGACTCAAC	392			
RESULT 2						
HSU62436	1876 bp	mRNA	PRI	11-JAN-1997		
LOCUS						
DEFINITION	Human nicotinic acetylcholine receptor alpha7 subunit precursor, mRNA, complete cds.					
ACCESSION	U62436					
NID	GI458119					
VERSION	U62436.1	GI:1458119				
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1876) Elliott,K.J., Ellis,S.B., Berckhan,K.J., Druetta,A., Chavez-Noriega,L.E., Johnson,E.C., Velicelcbl,G. and Harpold,M.M. Comparative structure of human neuronal alpha 2-alpha 7 and beta 2-beta 4 nicotinic acetylcholine receptor subunits and functional expression of the alpha 2, alpha 3, alpha 4, alpha 7, beta 2, and beta 4 subunits J. Mol. Neurosci. 7 (3), 217-228 (1996)					
AUTHORS	J. Mol. Neurosci. 7 (3), 217-228 (1996)					
TITLE	2 (bases 1 to 1876) Elliott,K.J. Direct Submission					

JOURNAL Submitted (28-JUN-1996) Kathryn J. Elliott, SIBR Neuroscience, Inc., 505 Coast Blvd. So., La Jolla, CA 92037, USA  
 FEATURES  
 source Location/Qualifiers  
 1..1876  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="KEalpha7.3"  
 /clone\_1ib="SIBR lambda gt11 library #3 (M. Williams)"  
 /cell\_line="IMR32"  
 /cell\_type="neuroblastoma"  
 1..72  
 1..138  
 73..1581  
 /codon\_start=1  
 /product="nicotinic acetylcholine receptor alpha7 subunit precursor"  
 /protein\_id="AAB40114.1"  
 /db\_xref="PDB:1A58120"  
 /db\_xref="GI:1458120"  
 /translation="MRCSPGVGVWLAIPASLHVSLSGEFQRLKYLEKYNPLERPV  
 ANDSOPLVYFSLIQIMDVDEKQVLTNTIMLQMSWTDHYLQNNVSEYPSGKTVRF  
 PQDQIKRTPILLIYNSADDERFDATFHTNVLVNSGHCQVLPQIFKRSCTIDVWPEPD  
 VQCKLRKSGMSYSGSLDQMEADISGVIYNGEWDLYIGPKRSERYECCKEPYP  
 DVFVTVMKRRLITYGLNLLIPVILSAALIVFLIPLADSGKISGLIDSLVYFM  
 LVAELMPTSDSVPLIAQIFASTIMTIVLSVYVTVYQYHHNPDGGMKMKLVFM  
 LLNMCWFLFLMKRPQDKYRPAQKQKRCSLASVMSVAPRPASNGLLYIGRGL  
 DGHVCPYDPSGVVCGRMACSPTHDEHLLHGGQPPGDPDLKILIEVRYINRRQ  
 DESEANQSEKRFANCAVVDRLCLMAFSVETIIIGILMSAPNVEAVSDFP"  
 139..1578  
 /product="nicotinic acetylcholine receptor alpha7 subunit"  
 1582..1876  
 3'UTR 369 a 553 c 531 g 423 t  
 BASE COUNT  
 ORIGIN  
 Query Match 18.4% Score 72; DB 10; Length 1876;  
 Best Local Similarity 100.0%; Pctd. No. 0.021; Indels 0; Gaps 0;  
 Matches 72; Conservative 0; Mismatches 0;  
 QY 321 GGCCCGAGGCGGAGGCCGGGGGACAGCGAGCTGTGAGCGCCGCGCTGCTGACACT 380  
 |||||||  
 DB 1 GGCCCGAGGCGGAGGCCGGGGGACAGCGAGCTGTGAGCGCGCCGCGCTGCTGACACT 60  
 |||||||  
 QY 381 CCGGAGCTCAAC 392  
 |||||||  
 DB 61 CCGGAGCTCAAC 72  
 |||||||  
 RESULT 3  
 AC007686  
 LOCUS AC007686 168919 bp DNA 30-MAY-1999  
 DEFINITION Homo sapiens chromosome 14 clone BAC 2289B16 map 14q24.3, LOW-PASS  
 SEQUENCE SAMPLING.  
 AC007686  
 94927297  
 AC007686.1 GI:4927297  
 HTG: HTGS\_PHASE0.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 168919)  
 2 (bases 1 to 168919)  
 Rowen, L., Madan, A., Qin, S., Abbasi, N., Dors, M., Dickhoff, R.,  
 James, R., Loretz, C., Lasky, S., Madan, A., Prescott, S., Ratcliffe, A.,  
 Shaffer, T. and Hood, L.  
 Sequencing of human chromosome 14q24.3 region  
 Unpublished  
 TITLE JOURNAL  
 JOURNAL  
 REFERENCE  
 REFERENCE  
 AUTHORS  
 TITLE JOURNAL  
 JOURNAL  
 Submitted (30-MAY-1999) Multimegabase Sequencing Center, University



## COMMENT

of Washington, PO BOX 357730, Seattle, WA 98195, USA  
\* NOTE: his record contains 192 individual.  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 965: contig of 965 bp in length  
\* gap of unknown length  
\* 966 1866: contig of 901 bp in length  
\* gap of unknown length  
\* 1867 2747: contig of 881 bp in length  
\* gap of unknown length  
\* 2748 3608: contig of 861 bp in length  
\* gap of unknown length  
\* 3609 4462: contig of 854 bp in length  
\* gap of unknown length  
\* 4463 5345: contig of 883 bp in length  
\* gap of unknown length  
\* 5346 6187: contig of 842 bp in length  
\* gap of unknown length  
\* 6188 7072: contig of 885 bp in length  
\* gap of unknown length  
\* 7073 7927: contig of 855 bp in length  
\* gap of unknown length  
\* 7928 8797: contig of 870 bp in length  
\* gap of unknown length  
\* 8798 9647: contig of 850 bp in length  
\* gap of unknown length  
\* 9648 10555: contig of 908 bp in length  
\* gap of unknown length  
\* 10556 11387: contig of 832 bp in length  
\* gap of unknown length  
\* 11388 12285: contig of 898 bp in length  
\* gap of unknown length  
\* 12286 13144: contig of 859 bp in length  
\* gap of unknown length  
\* 13145 14010: contig of 866 bp in length  
\* gap of unknown length  
\* 14011 14869: contig of 859 bp in length  
\* gap of unknown length  
\* 14870 15756: contig of 887 bp in length  
\* gap of unknown length  
\* 15757 16611: contig of 855 bp in length  
\* gap of unknown length  
\* 16612 17482: contig of 871 bp in length  
\* gap of unknown length  
\* 17483 18330: contig of 848 bp in length  
\* gap of unknown length  
\* 18331 19206: contig of 876 bp in length  
\* gap of unknown length  
\* 19207 20072: contig of 866 bp in length  
\* gap of unknown length  
\* 20073 20946: contig of 874 bp in length  
\* gap of unknown length  
\* 20947 21791: contig of 845 bp in length  
\* gap of unknown length  
\* 21792 23944: contig of 2153 bp in length  
\* gap of unknown length  
\* 23945 24793: contig of 849 bp in length  
\* gap of unknown length  
\* 24794 25663: contig of 870 bp in length  
\* gap of unknown length  
\* 25664 26504: contig of 841 bp in length  
\* gap of unknown length  
\* 26505 27381: contig of 877 bp in length  
\* gap of unknown length  
\* 27382 28254: contig of 873 bp in length  
\* gap of unknown length

28255 29126: gap of unknown length  
\* contig of 872 bp in length  
\* 29127 29963: gap of unknown length  
\* contig of 837 bp in length  
\* 29964 30822: gap of unknown length  
\* contig of 859 bp in length  
\* 30823 31680: gap of unknown length  
\* contig of 856 bp in length  
\* 31681 32538: gap of unknown length  
\* contig of 858 bp in length  
\* 32539 33390: gap of unknown length  
\* contig of 852 bp in length  
\* 33391 34255: gap of unknown length  
\* contig of 865 bp in length  
\* 34256 35105: gap of unknown length  
\* contig of 850 bp in length  
\* 35106 36030: gap of unknown length  
\* contig of 925 bp in length  
\* 36031 36869: gap of unknown length  
\* contig of 839 bp in length  
\* 36870 37734: gap of unknown length  
\* contig of 865 bp in length  
\* 37735 38584: gap of unknown length  
\* contig of 850 bp in length  
\* 38585 39448: gap of unknown length  
\* contig of 864 bp in length  
\* 39449 40283: gap of unknown length  
\* contig of 835 bp in length  
\* 40284 41169: gap of unknown length  
\* contig of 886 bp in length  
\* 41170 42014: gap of unknown length  
\* contig of 845 bp in length  
\* 42015 42927: gap of unknown length  
\* contig of 913 bp in length  
\* 42928 43788: gap of unknown length  
\* contig of 861 bp in length  
\* 43789 44650: gap of unknown length  
\* contig of 862 bp in length  
\* 44651 45491: gap of unknown length  
\* contig of 841 bp in length  
\* 45492 46383: gap of unknown length  
\* contig of 892 bp in length  
\* 46384 47232: gap of unknown length  
\* contig of 849 bp in length  
\* 47233 48096: gap of unknown length  
\* contig of 864 bp in length  
\* 48097 48951: gap of unknown length  
\* contig of 855 bp in length  
\* 48952 49831: gap of unknown length  
\* contig of 880 bp in length  
\* 49832 50676: gap of unknown length  
\* contig of 845 bp in length  
\* 50677 51534: gap of unknown length  
\* contig of 856 bp in length  
\* 51535 52388: gap of unknown length  
\* contig of 854 bp in length  
\* 52389 53266: gap of unknown length  
\* contig of 878 bp in length  
\* 53267 54108: gap of unknown length  
\* contig of 842 bp in length  
\* 54109 55136: gap of unknown length  
\* contig of 1028 bp in length  
\* 55137 55986: gap of unknown length  
\* contig of 850 bp in length  
\* 55987 56874: gap of unknown length  
\* contig of 888 bp in length  
\* 56875 57721: gap of unknown length  
\* contig of 847 bp in length  
\* 57722 58587: gap of unknown length  
\* contig of 866 bp in length  
\* 58588 59427: gap of unknown length  
\* contig of 840 bp in length  
\* gap of unknown length

[illegible]

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 2087)  
AUTHORS Peng, X., Katz, M., Gerzanich, V., Anand, R. and Lindstrom, J.  
TITLE Human alpha 7 acetylcholine receptor: cloning of the alpha 7  
subunit from the SH-SY5Y cell line and determination of  
pharmacological properties of native receptors and functional alpha  
7 homomers expressed in Xenopus oocytes  
Mol. Pharmacol. 45 (3), 546-554 (1994)  
JOURNAL MEDLINE 94195283  
REFERENCE 2 (bases 1 to 2087)  
AUTHORS Katz, M.  
TITLE Direct Submission  
JOURNAL Submitted (04-FEB-1993) M. Katz, Univ of Pennsylvania School of  
Medicine, Dept of Neuroscience, 36th & Hamilton Walk, Rm 235  
Stemmler Hall, Philadelphia, PA 19104, USA  
FEATURES  
source  
1. 2087  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="brain"  
/cell\_type="neuroblastoma"  
/clone\_lib="SHSY-5Y"  
/clone\_lib="lambda zap II"  
/clone="SHSY3"  
104. 175  
/product="neuronal nicotinic acetylcholine receptor  
alpha-7 subunit"  
104. 1612  
/codon\_start=1  
/product="neuronal nicotinic acetylcholine receptor  
alpha-7 subunit"  
/protein\_id="CAA49778.1"  
/db\_xref="PID:9496607"  
/db\_xref="GI:496607"  
/translation="MCSFGVWLGSLASLHVSLSQGEFQRLKLYELKYNPLERPV  
ANDSPLTYFSLSLQIMDVDEKNOVLTNTLMQSWTDHYLQNNVSEYPCVKTVE  
PDGQIMKPDILYNSADERFDATFTNTVYNSGHCQYLPPIFKSCYIDRWPEFD  
VOHCKLFGSNVSGMSLDLQOEADISGIVNGEMDLVGIPIKSEKSEYRECKEEDP  
DVTFTVTRRRRLTYGLNLIPICVLSALALVPLIPDSCGKISLGLTYLISLVYM  
LLVATIMPATSDSPLIAQNPASTMTIVGLSVVTVYTLQTHHDPDGKMKWTRVI  
LLNRCAMFLRMKRPGEDEKVPACQKHQRCSLAVEMSAVGPSPASNGSLYIGRGL  
DGVHCVPTPDSCVVCGRMACSEPTDHEHLHGQAPPEGDPDLKILEEYRILNFRRCQ  
DESEAVCESEMKFACVVDRLCLMAFSVFTICTIGILMSAPRFEAVSKDFA"  
176. 1609  
BASE COUNT 441 a 601 c 573 g 472 t  
IGIN  
mat\_peptide  
Query Match 17.8%; Score 69.8; DB 10; Length 2087;  
Best Local Similarity 95.9%; Pred. No. 0.041; 2; Indels 2; Gaps 2;  
Matches 93; Conservative 0; Mismatches 2; Indels 2; Gaps 2;  
QY 296 GAGCCGAGCGGCGAGTCTCTGTGCGGCGAGCGCCCGGCGAGCGAGAGC 355  
DB 9 GAGCCGAGCGGCGAGTCTCTGTGCGGCGAGCG-CAAGGCGGCGAGCA-CGAGAGC 66  
QY 356 TGGAGCGCGCGGCTGCTGACGACTCCGGGACTCAAC 392  
DB 67 TGGAGCGCGCGGCTGCTGACGACTCCGGGACTCAAC 103  
RESULT 6  
LOCUS HS21EA 1072 bp DNA VRL 15-MAR-1990  
DEFINITION Herpes simplex virus type 2 immediate-early (IE5) protein mRNA, 5'  
end.  
ACCESSION M29385  
NID 9330283  
VERSION M29385.1 GI:330283  
KEYWORDS

SOURCE  
ORGANISM Herpes simplex virus type 2 (strain HGS Timbury, 1971) DNA, passed  
in baby hamster kidney clone 21 (C13) cells.  
REFERENCE 1 (bases 1 to 1072)  
AUTHORS Whitton, J. L. and Clements, J. B.  
TITLE The junctions between the repetitive and the short unique sequences  
of the herpes simplex virus genome are determined by the  
polypeptide-coding regions of two spliced immediate-early mRNAs  
J. Gen. Virol. 65, 451-466 (1984)  
JOURNAL MEDLINE 84137573  
FEATURES  
source  
1. 1072  
/organism="human herpesvirus 2"  
/db\_xref="taxon:10310"  
prim\_transcript 95. >1072  
/note="IE-5 mRNA and intron"  
344. 866  
/note="IE-5 intron"  
897. >1072  
/note="immediate early protein 5"  
/codon\_start=1  
/protein\_id="AAA45849.1"  
/db\_xref="PID:9330284"  
/db\_xref="GI:330284"  
/translation="MSNALKRTTDMPSDSSRCHRTHTYGVCAEIKHREDEARNTAV  
TDPEPLICPDVPR"  
BASE COUNT 132 a 322 c 543 g 75 t  
ORIGIN  
Query Match 17.3%; Score 68; DB 17; Length 1072;  
Best Local Similarity 50.6%; Pred. No. 0.091;  
Matches 164; Conservative 0; Mismatches 160; Indels 0; Gaps 0;  
QY 37 GCCCTGAGCTCGTCCAGAGGCGGAGAGCCGCTGCTGAGAGTGGGGGTGA 96  
DB 366 GCGGGGAGGAGCGGCGGAGGAGCGGCGGAGGAGCGGCGGAGGAGCGGCGG 425  
QY 97 GGTCCCGGAGCGTACCGAGCGGCGGAGTACCTCCGCTCACCTCGAGTTC 156  
DB 426 GACGGGCGGCGGAGCGGCGGCGGAGCGGCGGCGGAGCGGCGGCGGCGG 485  
QY 157 COTGGGTGCGCGGAGAGCGTGGCGCGGCGGAGAGGATGGCGGCGGCGGCG 216  
DB 486 GCGGGGCGGCGGAGAGCGGCGGCGGAGCGGCGGCGGAGCGGCGGCGGCG 545  
QY 217 GGGGCGGCGCTCTACAGTGAAGCGCGGCGGCGGCGGCGGCGGCGCGCC 276  
DB 546 CGGGGAGAGCGGCGGAGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGCG 605  
QY 277 CGGCTCTTAAAGCGCGGAGCGGCGGAGTGGCTCTGTGCGCGGAGCGGCG 336  
DB 606 GGGGAGCGGCGGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGG 665  
QY 337 CCGGCGAGAGCGGAGAGTGAAG 360  
DB 666 CCGGGGAGCGGCGGCGGCGGCGG 689  
RESULT 7  
LOCUS HS21E 1560 bp DNA VRL 15-MAR-1990  
DEFINITION Herpes simplex virus type 2 immediate-early (IE4) protein mRNA, 5'  
end.  
ACCESSION M29384  
NID 9330281  
VERSION M29384.1 GI:330281  
KEYWORDS  
SOURCE Herpes simplex virus type 2 (strain HGS Timbury, 1971) DNA, passed  
in baby hamster kidney clone 21 (C13) cells.  
ORGANISM human herpesvirus 2

Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Alphaherpesvirinae; Simplexvirus.

REFERENCE 1 (bases 1 to 1560)

AUTHORS Whittton,J.L. and Clements,J.B.  
TITLE The junctions between the repetitive and the short unique sequences of the herpes simplex virus genome are determined by the polypeptide-coding regions of two spliced immediate-early mRNAs

JOURNAL J. Gen. Virol. 65, 451-466 (1984)

JOURNAL 84137573

FEATURES

source Location/Qualifiers

1. .1560 /organism="human herpesvirus 2"

/db\_xref="taxon:10310"

prim\_transcript 95. >1560 /note="IE-4 mRNA and Intron"

Intron 344. .906 /note="IE-4 Intron"

CDS 969. >1560 /note="Immediate early protein 4"

/codon\_start=1 /protein\_id="AA45848.1"

/db\_xref="GI:555152"

/translation="MADIPDPDLPALPTTPANNAHPSPPLPDSRRRRPVPSPSESGX  
PDIESSSTSESEDEVDLGRRRPRLGRRFLDLSAETSTSESGTSDDD  
DDASDGMVDPKRRKRRPRLNRLTSLSPDRAGVFFPVRMSDRIRAAQQAQAPAO  
SSDRAHAPRRSRAQOMRSGAMTLDLHYIRQCVDN"

BASE COUNT 222 a 480 c 714 g 144 t

ORIGIN

Query Match 16.8%; Score 65.8; DB 17; Length 1560;  
Best Local Similarity 51.3%; Pred. No. 0.16; Mismatches 167; Indels 1; Gaps 1;

Matches 177; Conservative 0; Mismatches 167; Indels 1; Gaps 1;

25 TGGCTTGGGAGAGCCCTGGCTGCTGAGAGGCGGAGAGCCGAGAGCCGCTCGGTGAG 84

350 TGACCGGCG 409

85 ACTGGGGGTGAGAGTTCGCCGCGAGCGCTACCGAGCGCGGAGAGTCCCGCTACACCTC 144

410 ACGGCGCGGCGGAGAGCG 469

145 GGGCTGCAAGTTCCTGGTGGCGCGCGCGAGAGCGCTGCGCGCGCGCGCGCGCGCGCG 204

470 GGGGCGCGGCG 529

205 GGGGAGCGGCGCGCGCGCGCGCGCGCTGTCAGTGAAGAGCGCGCGCGCGCGCGCGCG 264

529 GGGGCGCGGCGGAGCG 588

265 GGGGCGCGGCGCGCGCGCGCTCTTAAGCGCGCGCGAGCGCGCGAGTGCCTCTGTGCGC 324

589 GGGGCGCGGCG 648

325 GCAGGCGCGAGCGCGCGCGCGAGCGCGAGAGTGAAGCGCGCGCGCGCGCGCGCG 369

649 GGGAGCGGCG 693

RESULT 8  
LOCUS HSGG1 43058 bp DNA PRI 19-MAR-1997  
DEFINITION Human DNA sequence from cosmid GGI from a contig from the tip of  
the short arm of chromosome 16, spanning 2Mb of 16p13.3 Contains  
alpha and zeta globin genes and ESTs.

ACCESSION 284721  
VERSION 1817575  
KEYWORDS 16p13.3; alpha-globin; globin; zeta-globin.

SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 43058)

AUTHORS Flint,J. and Higgs,D.R.

TITLE Direct Submission

JOURNAL Submitted (28-JAN-1997) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1RO, UK. E-mail enquiries: humphdesanger.ac.uk

IMPORTANT: This sequence is the entire insert of clone GGI. This  
clone was sequenced at the Institute of Molecular Medicine. The true right  
true left end of clone GGI is at 1 in this sequence. The true right  
end of clone GGI is at 43058.

16p. Higgs D.R., Flint J. unpublished. MRC Molecular Haematology  
Unit, Institute of Molecular Medicine, Oxford.

GCI came from the Los Alamos, flow-sorted human Chromosome 16  
library.

FEATURES

source

Location/Qualifiers

1. .43058

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="16"

/map="16p13.3"

/clone="GGI"

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

```

repeat_region
/note="AluY repeat: matches 13. .294 of consensus"
10382. .10428
/note="AluX/g repeat: matches 245. .291 of consensus;
incomplete repeat"
repeat_region
/note="L1MB8 repeat: matches 920. .859 of consensus"
10652. .10713
/note="L1MB8 repeat: matches 920. .859 of consensus"
10714. .10966
/note="AluSp repeat: matches 1. .255 of consensus;
incomplete repeat"
repeat_region
11022. .11168
/note="AluX repeat: matches 1. .145 of consensus;
incomplete repeat"
repeat_region
/note="AluSp repeat: matches 1. .296 of consensus"
11466. .11635
/note="AluX repeat: matches 133. .302 of consensus;
incomplete repeat"
repeat_region
/note="L1 repeat: matches 4729. .5317 of consensus"
11730. .12317
/note="L1 repeat: matches 4729. .5317 of consensus"
12321. .12616
/note="AluY repeat: matches 1. .296 of consensus"
12633. .12761
/note="AluSp repeat: matches 1. .131 of consensus;
incomplete repeat"
repeat_region
12779. .13073
/note="AluSp repeat: matches 1. .297 of consensus"
13011.13773. .13867,14755. .14959,15135. .15263)
/codon_start=1
/product="zeta-globin 1"
/protein_id="CAB06552.1"
/db_xref="PID:3300370"
/db_xref="PID:91817576"
/db_xref="GI:1817576"
/db_xref="GI:1817576"
/translacion="MSLRTERTIIVSMWAKISQADIGNETLRLFLSHPOKTYE
PHEFDHPGSAQIRANGSKVYANGDAVSDIDGALSKELEATVILTRDPVNFKLL
SHCLVLTAAARRPADFTLEAHAMKFLSVSVLTXEYR"
13923. .14090
/note="12 copies of 14 mer 89 & conserved"
14618. .15730
/note="Putative Cpg Island"
14964. .15071
/note="3 copies of 36 mer 83 & conserved"
15747. .15875
/note="MIR2 repeat: matches 145. .10 of consensus"
16241. .16541
/note="AluY repeat: matches 301. .1 of consensus"
16892. .17027
/note="AluJo repeat: matches 1. .136 of consensus;
incomplete repeat"
repeat_region
17029. .17326
/note="AluY repeat: matches 1. .297 of consensus"
17368. .17506
/note="AluJo repeat: matches 10. .151 of consensus;
incomplete repeat"
repeat_region
17945. .18086
/note="AluJo repeat: matches 153. .294 of consensus;
incomplete repeat"
repeat_region
18106. .18426
/note="AluJo repeat: matches 1. .301 of consensus"
18715. .18833
/note="L1MB7 repeat: matches 744. .865 of consensus"
18834. .19130
/note="AluX repeat: matches 1. .296 of consensus"
19186. .19449
/note="MER42C repeat: matches 1506. .1230 of consensus"
19431. .19593
/note="HUMAR1 repeat: matches 1221. .1058 of consensus"
19601. .19838
/note="AluSp repeat: matches 300. .63 of consensus;
incomplete repeat"
repeat_region
19840. .19983
/note="AluSp repeat: matches 154. .300 of consensus;

```

```

repeat_region    20709..20856
                  incomplete repeat"
                  /note="4 copies of 37 mer 93 % conserved"
repeat_region    20861..20952
                  /note="2 copies of 36 mer 99 % conserved"
repeat_region    20919..21029
                  /note="3 copies of 37 mer 97 % conserved"
repeat_region    20996..21067
                  /note="2 copies of 36 mer 99 % conserved"
repeat_region    21039..21186
                  /note="4 copies of 37 mer 95 % conserved"
repeat_region    21216..21323
                  /note="3 copies of 36 mer 97 % conserved"
repeat_region    21325..21396
                  /note="2 copies of 36 mer 97 % conserved"
repeat_region    21360..21433
                  /note="2 copies of 37 mer 97 % conserved"
repeat_region    21398..21469
                  /note="2 copies of 36 mer 97 % conserved"
repeat_region    21471..21542
                  /note="2 copies of 36 mer 99 % conserved"
repeat_region    21546..21617
                  /note="2 copies of 36 mer 96 % conserved"
repeat_region    21579..21726
                  /note="4 copies of 37 mer 94 % conserved"
repeat_region    22591..22661
                  /note="AluX repeat: matches 302..232 of consensus;
incomplete repeat"
repeat_region    22766..23070
                  /note="AluX repeat: matches 1..302 of consensus"
CDS              join(23985..24079,25345..25549,25891..26019)
                  /codon_start=1
                  /pseudo
                  /product="zeta-globin 2"
                  /db_xref="PID:es00371"
                  24135..24680
                  /note="39 copies of 14 mer 90 % conserved"
misc_feature      25208..25784
                  /note="Putative Cpg island"
                  <26871..>27290
                  /note="match: multiple ESTs; similar to HEMOGLOBIN
ALPHA-B; match: 5' EST H57023 clone 204625; match: 5' EST
R93558 clone 196372; match: 3' EST H67154 clone 211156"
prim_transcript

```

Query Match	16.1%;	Score 63.2;	DB 10;	Length 43058;
Best Local Similarity	52.7%;	Pred. No. 0.12;		
Matches 137;	Conservative	0;	Mismatches 123;	Indels 0;
				Gaps 0

OY	24	CTGGGCTTTGGGGAGCGCCCTGGGCTGTGGCCACAGAAAGGCAGAGGCCGAGAACCCTGTGGTGA	83
Db	25599	C GG G C C G G G C C G G G G C C C G G G G G T T C G C G G G G C C G G G G T G G G T G C C G G G C G	25648
OY	84	GACTGGGGGTGAAGGTGGCCGAGAGCCTAACCAAGCCGCGAATATTCTTCGCTTCACACT	143
Db	25649	G G G C G G G G T T C C G C G G G C G G G G G G G G G G G G G G G G G G C C G G G C C G G G C C G G G C	25708
OY	144	CGGGCTTGCAATTCCCTGGGTGTGGCCCGACAGACGTGTGGCCCCGGGGCTGTGAAGGATGTGGCGGG	203
Db	25709	G G G G C G G G G C C G G G G G G A G G G C T T G G C G G G G C G G G G C G G G G C G G G G C C G G G	25768
OY	204	CGGGGACGGGGCGGGGGCGGGGCGTGTCTCACATGTGAAGAGCGGCGCGGGCGGGGCG	263
Db	25769	C C G G G G C G G G G T C G C G G G G G G G T C G C G G G G C G G G G C G C G G G G C G G G G C G G G G T	25828
OY	264	GGGGGCGGCGCCCGGCGCTTC	283
Db	25829	G G G G T C G C G G G G C G G G G C C C	25848

RESULT	9
HUMHB3	
LOCUS	HUMHB3 2685 bp DNA
DEFINITION	Human alpha globin gene cluster on chromosome 16: psfeta
	18-SEP-1997

Accession J00184  
 NID 9183792  
 VERSION J00184.1 GI:183792  
 KEYWORDS gene duplication; pseudogene; repeat region; zeta-globin.  
 SEGMENT 3 of 4  
 SOURCE Homo sapiens DNA.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 2885)  
 PROUDFOOT,N.J., GILL,A. and MANIATIS,T.  
 The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene  
 Cell 31 (3 Pt 2), 553-563 (1982)  
 JOURNAL 83129370  
 MEDLINE  
 COMMENT The human alpha globin gene cluster, located on the short arm of chromosome 16, spans about 30 kb and includes the following five loci:  
 5'-zeta-pseudozeta-pseudozeta-1-alpha-2-alpha-1-3'. The gene sequence shown below is considered a pseudogene because of the terminating codon at base 429. There are only two additional single-base differences between the coding sequences of the zeta and pseudozeta genes, although there are conspicuous differences in the 5' and 3' flanking regions. Moreover the first intron of this locus demonstrates length variation, apparently stemming from the repeat sequence 'acagtgagagag' which is also found in the zeta gene and, in a similar form, in the 5' flank of the human insulin gene. The repeat sequence 'cgagg' is characteristic of the second introns of the zeta genes [1].  
 Typical promoter elements 'ccat' and 'tata' are found at bases 291 and 326.

# FEATURES

Source 1.2685  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="16p13.3"  
 356..505  
 /gene="phbz"  
 /number=1  
 join(411..505,1767..1971,2313..2441)  
 /gene="phbz"  
 /pseudo  
 /codon\_start=1  
 join(411..505,1767..1971,2313..2441)  
 /gene="phbz"  
 mutation 429  
 /gene="phbz"  
 /note="t in phbz; g in hbz (see <humhbal>)"  
 506..1766  
 /gene="phbz"  
 /number=1  
 repeat\_region 565..578  
 /note="repeat sequence copy 1"  
 1767..1971  
 /gene="phbz"  
 intron 1972..2312  
 /gene="phbz"  
 repeat\_region 1977..1981  
 /note="repeat sequence, copy 1"  
 2313..2541  
 /gene="phbz"  
 /number=3  
 2521..2526  
 /note="phbz mRNA polyadenylation signal"  
 BASE COUNT 454 a 720 c 1155 g 356 t  
 ORIGIN About 1 kb after segment 2.

Query Match 16.1%; Score 63.2; DB 10; Length 2685;  
 Best Local Similarity 52.7%; Pred. No. 0.31;

Matches 137; Conservative 0; Mismatches 123; Indels 0; Gaps 0;  
 QY 24 CTGGCCCTTGGCAACCCCTGCGCCAGAGCGCCGAGACCCGCGTGGTGA 83  
 Db 2011 CCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG 2070  
 QY 84 GACHTGGGGTGGAGTGCCTCCGAGACCGCGGAGTAGTACTCCGCTCACACT 143  
 Db 2071 GGGCGGGGTCTCCGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG 2130  
 QY 144 CGGCTCAGTTCCTGAGTGGCCGCCGAGACGCTGGCCCGGCTGAGAGATGCGGGC 203  
 Db 2131 GGGCGGGCGGGCGGGCGGGAGAGGGCTGGCGGGCGGGCGGGCGGGCGGGCG 2190  
 QY 204 CGGGAGCGGGCGGGCGGGCGGGCTGCTGACGTGAGAGCGCCCGGGCGGGCGG 263  
 Db 2191 CCGGGCGGGCTCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGGT 2250  
 QY 264 GGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGGT 2250  
 Db 2251 GGGGTGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGGT 2270  
 RESULT 10  
 LOCUS HSV2HG52 154746 bp DNA VRL 04-DEC-1998  
 DEFINITION Herpes simplex virus type 2 (strain HG52), complete genome.  
 ACCESSION 286099  
 NID 91869820  
 VERSION 286099.1 GI:1869820  
 KEYWORDS capsid associated tegument protein; capsid protein; complete genome; deoxyribonuclease; deoxyribonuclease; DNA polymerase; envelope protein; glycoprotein B; glycoprotein C; glycoprotein D; glycoprotein E; glycoprotein G; glycoprotein H; glycoprotein I; glycoprotein J; glycoprotein K; glycoprotein L; glycoprotein M; host shut-off factor; integral membrane protein; major capsid protein; membrane-associated phosphoprotein; minor capsid protein; neurovirulence factor; protease; protein kinase; ribonucleotide reductase large subunit; ribonucleotide reductase small subunit; R1 gene; R2 gene; R3 gene; tegument protein; thymidine kinase; U1 gene; U10 gene; U11 gene; U12 gene; U13 gene; U14 gene; U15 gene; U16 gene; U17 gene; U18 gene; U19 gene; U20 gene; U21 gene; U22 gene; U23 gene; U24 gene; U25 gene; U26 gene; U26.5 gene; U27 gene; U28 gene; U29 gene; U30 gene; U31 gene; U32 gene; U33 gene; U34 gene; U35 gene; U36 gene; U37 gene; U38 gene; U39 gene; U40 gene; U41 gene; U42 gene; U43 gene; U44 gene; U45 gene; U46 gene; U47 gene; U48 gene; U49 gene; U50 gene; U51 gene; U52 gene; U53 gene; U54 gene; U55 gene; U56 gene; U57 gene; U58 gene; U59 gene; U60 gene; U61 gene; U62 gene; U63 gene; U64 gene; U65 gene; U66 gene; U67 gene; U68 gene; U69 gene; U70 gene; U71 gene; U72 gene; U73 gene; U74 gene; U75 gene; U76 gene; U77 gene; U78 gene; U79 gene; U80 gene; U81 gene; U82 gene; U83 gene; U84 gene; U85 gene; U86 gene; U87 gene; U88 gene; U89 gene; U90 gene; U91 gene; U92 gene; U93 gene; U94 gene; U95 gene; U96 gene; U97 gene; U98 gene; U99 gene; U100 gene; U101 gene; U102 gene; U103 gene; U104 gene; U105 gene; U106 gene; U107 gene; U108 gene; U109 gene; U110 gene; U111 gene; U112 gene; U113 gene; U114 gene; U115 gene; U116 gene; U117 gene; U118 gene; U119 gene; U120 gene; U121 gene; U122 gene; U123 gene; U124 gene; U125 gene; U126 gene; U126.5 gene; U127 gene; U128 gene; U129 gene; U130 gene; U131 gene; U132 gene; U133 gene; U134 gene; U135 gene; U136 gene; U137 gene; U138 gene; U139 gene; U140 gene; U141 gene; U142 gene; U143 gene; U144 gene; U145 gene; U146 gene; U147 gene; U148 gene; U149 gene; U150 gene; U151 gene; U152 gene; U153 gene; U154 gene; U155 gene; U156 gene; U157 gene; U158 gene; U159 gene; U160 gene; U161 gene; U162 gene; U163 gene; U164 gene; U165 gene; U166 gene; U167 gene; U168 gene; U169 gene; U170 gene; U171 gene; U172 gene; U173 gene; U174 gene; U175 gene; U176 gene; U177 gene; U178 gene; U179 gene; U180 gene; U181 gene; U182 gene; U183 gene; U184 gene; U185 gene; U186 gene; U187 gene; U188 gene; U189 gene; U190 gene; U191 gene; U192 gene; U193 gene; U194 gene; U195 gene; U196 gene; U197 gene; U198 gene; U199 gene; U200 gene; U201 gene; U202 gene; U203 gene; U204 gene; U205 gene; U206 gene; U207 gene; U208 gene; U209 gene; U210 gene; U211 gene; U212 gene; U213 gene; U214 gene; U215 gene; U216 gene; U217 gene; U218 gene; U219 gene; U220 gene; U221 gene; U222 gene; U223 gene; U224 gene; U225 gene; U226 gene; U226.5 gene; U227 gene; U228 gene; U229 gene; U230 gene; U231 gene; U232 gene; U233 gene; U234 gene; U235 gene; U236 gene; U237 gene; U238 gene; U239 gene; U240 gene; U241 gene; U242 gene; U243 gene; U244 gene; U245 gene; U246 gene; U247 gene; U248 gene; U249 gene; U250 gene; U251 gene; U252 gene; U253 gene; U254 gene; U255 gene; U256 gene; U257 gene; U258 gene; U259 gene; U260 gene; U261 gene; U262 gene; U263 gene; U264 gene; U265 gene; U266 gene; U267 gene; U268 gene; U269 gene; U270 gene; U271 gene; U272 gene; U273 gene; U274 gene; U275 gene; U276 gene; U277 gene; U278 gene; U279 gene; U280 gene; U281 gene; U282 gene; U283 gene; U284 gene; U285 gene; U286 gene; U287 gene; U288 gene; U289 gene; U290 gene; U291 gene; U292 gene; U293 gene; U294 gene; U295 gene; U296 gene; U297 gene; U298 gene; U299 gene; U300 gene; U301 gene; U302 gene; U303 gene; U304 gene; U305 gene; U306 gene; U307 gene; U308 gene; U309 gene; U310 gene; U311 gene; U312 gene; U313 gene; U314 gene; U315 gene; U316 gene; U317 gene; U318 gene; U319 gene; U320 gene; U321 gene; U322 gene; U323 gene; U324 gene; U325 gene; U326 gene; U326.5 gene; U327 gene; U328 gene; U329 gene; U330 gene; U331 gene; U332 gene; U333 gene; U334 gene; U335 gene; U336 gene; U337 gene; U338 gene; U339 gene; U340 gene; U341 gene; U342 gene; U343 gene; U344 gene; U345 gene; U346 gene; U347 gene; U348 gene; U349 gene; U350 gene; U351 gene; U352 gene; U353 gene; U354 gene; U355 gene; U356 gene; U357 gene; U358 gene; U359 gene; U360 gene; U361 gene; U362 gene; U363 gene; U364 gene; U365 gene; U366 gene; U367 gene; U368 gene; U369 gene; U370 gene; U371 gene; U372 gene; U373 gene; U374 gene; U375 gene; U376 gene; U377 gene; U378 gene; U379 gene; U380 gene; U381 gene; U382 gene; U383 gene; U384 gene; U385 gene; U386 gene; U387 gene; U388 gene; U389 gene; U390 gene; U391 gene; U392 gene; U393 gene; U394 gene; U395 gene; U396 gene; U397 gene; U398 gene; U399 gene; U400 gene; U401 gene; U402 gene; U403 gene; U404 gene; U405 gene; U406 gene; U407 gene; U408 gene; U409 gene; U410 gene; U411 gene; U412 gene; U413 gene; U414 gene; U415 gene; U416 gene; U417 gene; U418 gene; U419 gene; U420 gene; U421 gene; U422 gene; U423 gene; U424 gene; U425 gene; U426 gene; U426.5 gene; U427 gene; U428 gene; U429 gene; U430 gene; U431 gene; U432 gene; U433 gene; U434 gene; U435 gene; U436 gene; U437 gene; U438 gene; U439 gene; U440 gene; U441 gene; U442 gene; U443 gene; U444 gene; U445 gene; U446 gene; U447 gene; U448 gene; U449 gene; U450 gene; U451 gene; U452 gene; U453 gene; U454 gene; U455 gene; U456 gene; U457 gene; U458 gene; U459 gene; U460 gene; U461 gene; U462 gene; U463 gene; U464 gene; U465 gene; U466 gene; U467 gene; U468 gene; U469 gene; U470 gene; U471 gene; U472 gene; U473 gene; U474 gene; U475 gene; U476 gene; U477 gene; U478 gene; U479 gene; U480 gene; U481 gene; U482 gene; U483 gene; U484 gene; U485 gene; U486 gene; U487 gene; U488 gene; U489 gene; U490 gene; U491 gene; U492 gene; U493 gene; U494 gene; U495 gene; U496 gene; U497 gene; U498 gene; U499 gene; U500 gene; U501 gene; U502 gene; U503 gene; U504 gene; U505 gene; U506 gene; U507 gene; U508 gene; U509 gene; U510 gene; U511 gene; U512 gene; U513 gene; U514 gene; U515 gene; U516 gene; U517 gene; U518 gene; U519 gene; U520 gene; U521 gene; U522 gene; U523 gene; U524 gene; U525 gene; U526 gene; U527 gene; U528 gene; U529 gene; U530 gene; U531 gene; U532 gene; U533 gene; U534 gene; U535 gene; U536 gene; U537 gene; U538 gene; U539 gene; U540 gene; U541 gene; U542 gene; U543 gene; U544 gene; U545 gene; U546 gene; U547 gene; U548 gene; U549 gene; U550 gene; U551 gene; U552 gene; U553 gene; U554 gene; U555 gene; U556 gene; U557 gene; U558 gene; U559 gene; U560 gene; U561 gene; U562 gene; U563 gene; U564 gene; U565 gene; U566 gene; U567 gene; U568 gene; U569 gene; U570 gene; U571 gene; U572 gene; U573 gene; U574 gene; U575 gene; U576 gene; U577 gene; U578 gene; U579 gene; U580 gene; U581 gene; U582 gene; U583 gene; U584 gene; U585 gene; U586 gene; U587 gene; U588 gene; U589 gene; U590 gene; U591 gene; U592 gene; U593 gene; U594 gene; U595 gene; U596 gene; U597 gene; U598 gene; U599 gene; U600 gene; U601 gene; U602 gene; U603 gene; U604 gene; U605 gene; U606 gene; U607 gene; U608 gene; U609 gene; U610 gene; U611 gene; U612 gene; U613 gene; U614 gene; U615 gene; U616 gene; U617 gene; U618 gene; U619 gene; U620 gene; U621 gene; U622 gene; U623 gene; U624 gene; U625 gene; U626 gene; U626.5 gene; U627 gene; U628 gene; U629 gene; U630 gene; U631 gene; U632 gene; U633 gene; U634 gene; U635 gene; U636 gene; U637 gene; U638 gene; U639 gene; U640 gene; U641 gene; U642 gene; U643 gene; U644 gene; U645 gene; U646 gene; U647 gene; U648 gene; U649 gene; U650 gene; U651 gene; U652 gene; U653 gene; U654 gene; U655 gene; U656 gene; U657 gene; U658 gene; U659 gene; U660 gene; U661 gene; U662 gene; U663 gene; U664 gene; U665 gene; U666 gene; U667 gene; U668 gene; U669 gene; U670 gene; U671 gene; U672 gene; U673 gene; U674 gene; U675 gene; U676 gene; U677 gene; U678 gene; U679 gene; U680 gene; U681 gene; U682 gene; U683 gene; U684 gene; U685 gene; U686 gene; U687 gene; U688 gene; U689 gene; U690 gene; U691 gene; U692 gene; U693 gene; U694 gene; U695 gene; U696 gene; U697 gene; U698 gene; U699 gene; U700 gene; U701 gene; U702 gene; U703 gene; U704 gene; U705 gene; U706 gene; U707 gene; U708 gene; U709 gene; U710 gene; U711 gene; U712 gene; U713 gene; U714 gene; U715 gene; U716 gene; U717 gene; U718 gene; U719 gene; U720 gene; U721 gene; U722 gene; U723 gene; U724 gene; U725 gene; U726 gene; U726.5 gene; U727 gene; U728 gene; U729 gene; U730 gene; U731 gene; U732 gene; U733 gene; U734 gene; U735 gene; U736 gene; U737 gene; U738 gene; U739 gene; U740 gene; U741 gene; U742 gene; U743 gene; U744 gene; U745 gene; U746 gene; U747 gene; U748 gene; U749 gene; U750 gene; U751 gene; U752 gene; U753 gene; U754 gene; U755 gene; U756 gene; U757 gene; U758 gene; U759 gene; U760 gene; U761 gene; U762 gene; U763 gene; U764 gene; U765 gene; U766 gene; U767 gene; U768 gene; U769 gene; U770 gene; U771 gene; U772 gene; U773 gene; U774 gene; U775 gene; U776 gene; U777 gene; U778 gene; U779 gene; U780 gene; U781 gene; U782 gene; U783 gene; U784 gene; U785 gene; U786 gene; U787 gene; U788 gene; U789 gene; U790 gene; U791 gene; U792 gene; U793 gene; U794 gene; U795 gene; U796 gene; U797 gene; U798 gene; U799 gene; U800 gene; U801 gene; U802 gene; U803 gene; U804 gene; U805 gene; U806 gene; U807 gene; U808 gene; U809 gene; U810 gene; U811 gene; U812 gene; U813 gene; U814 gene; U815 gene; U816 gene; U817 gene; U818 gene; U819 gene; U820 gene; U821 gene; U822 gene; U823 gene; U824 gene; U825 gene; U826 gene; U826.5 gene; U827 gene; U828 gene; U829 gene; U830 gene; U831 gene; U832 gene; U833 gene; U834 gene; U835 gene; U836 gene; U837 gene; U838 gene; U839 gene; U840 gene; U841 gene; U842 gene; U843 gene; U844 gene; U845 gene; U846 gene; U847 gene; U848 gene; U849 gene; U850 gene; U851 gene; U852 gene; U853 gene; U854 gene; U855 gene; U856 gene; U857 gene; U858 gene; U859 gene; U860 gene; U861 gene; U862 gene; U863 gene; U864 gene; U865 gene; U866 gene; U867 gene; U868 gene; U869 gene; U870 gene; U871 gene; U872 gene; U873 gene; U874 gene; U875 gene; U876 gene; U877 gene; U878 gene; U879 gene; U880 gene; U881 gene; U882 gene; U883 gene; U884 gene; U885 gene; U886 gene; U887 gene; U888 gene; U889 gene; U890 gene; U891 gene; U892 gene; U893 gene; U894 gene; U895 gene; U896 gene; U897 gene; U898 gene; U899 gene; U900 gene; U901 gene; U902 gene; U903 gene; U904 gene; U905 gene; U906 gene; U907 gene; U908 gene; U909 gene; U910 gene; U911 gene; U912 gene; U913 gene; U914 gene; U915 gene; U916 gene; U917 gene; U918 gene; U919 gene; U920 gene; U921 gene; U922 gene; U923 gene; U924 gene; U925 gene; U926 gene; U926.5 gene; U927 gene; U928 gene; U929 gene; U930 gene; U931 gene; U932 gene; U933 gene; U934 gene; U935 gene; U936 gene; U937 gene; U938 gene; U939 gene; U940 gene; U941 gene; U942 gene; U943 gene; U944 gene; U945 gene; U946 gene; U947 gene; U948 gene; U949 gene; U950 gene; U951 gene; U952 gene; U953 gene; U954 gene; U955 gene; U956 gene; U957 gene; U958 gene; U959 gene; U960 gene; U961 gene; U962 gene; U963 gene; U964 gene; U965 gene; U966 gene; U967 gene; U968 gene; U969 gene; U970 gene; U971 gene; U972 gene; U973 gene; U974 gene; U975 gene; U976 gene; U977 gene; U978 gene; U979 gene; U980 gene; U981 gene; U982 gene; U983 gene; U984 gene; U985 gene; U986 gene; U987 gene; U988 gene; U989 gene; U990 gene; U991 gene; U992 gene; U993 gene; U994 gene; U995 gene; U996 gene; U997 gene; U998 gene; U999 gene; U1000 gene; U1001 gene; U1002 gene; U1003 gene; U1004 gene; U1005 gene; U1006 gene; U1007 gene; U1008 gene; U1009 gene; U1010 gene; U1011 gene; U1012 gene; U1013 gene; U1014 gene; U1015 gene; U1016 gene; U1017 gene; U1018 gene; U1019 gene; U1020 gene; U1021 gene; U1022 gene; U1023 gene; U1024 gene; U1025 gene; U1026 gene; U1026.5 gene; U1027 gene; U1028 gene; U1029 gene; U1030 gene; U1031 gene; U1032 gene; U1033 gene; U1034 gene; U1035 gene; U1036 gene; U1037 gene; U1038 gene; U1039 gene; U1040 gene; U1041 gene; U1042 gene; U1043 gene; U1044 gene; U1045 gene; U1046 gene; U1047 gene; U1048 gene; U1049 gene; U1050 gene; U1051 gene; U1052 gene; U1053 gene; U1054 gene; U1055 gene; U1056 gene; U1057 gene; U1058 gene; U1059 gene; U1060 gene; U1061 gene; U1062 gene; U1063 gene; U1064 gene; U1065 gene; U1066 gene; U1067 gene; U1068 gene; U1069 gene; U1070 gene; U1071 gene; U1072 gene; U1073 gene; U1074 gene; U1075 gene; U1076 gene; U1077 gene; U1078 gene; U1079 gene; U1080 gene; U1081 gene; U1082 gene; U1083 gene; U1084 gene; U1085 gene; U1086 gene; U1087 gene; U1088 gene; U1089 gene; U1090 gene; U1091 gene; U1092 gene; U1093 gene; U1094 gene; U1095 gene; U1096 gene; U1097 gene; U1098 gene; U1099 gene; U1100 gene; U1101 gene; U1102 gene; U1103 gene; U1104 gene; U1105 gene; U1106 gene; U1107 gene; U1108 gene; U1109 gene; U1110 gene; U1111 gene; U1112 gene; U1113 gene; U1114 gene; U1115 gene; U1116 gene; U1117 gene; U1118 gene; U1119 gene; U1120 gene; U1121 gene; U1122 gene; U1123 gene; U1124 gene; U1125 gene; U1126 gene; U1126.5 gene; U1127 gene; U1128 gene; U1129 gene; U1130 gene; U1131 gene; U1132 gene; U1133 gene; U1134 gene; U1135 gene; U1136 gene; U1137 gene; U1138 gene; U1139 gene; U1140 gene; U1141 gene; U1142 gene; U1143 gene; U1144 gene; U1145 gene; U1146 gene; U1147 gene; U1148 gene; U1149 gene; U1150 gene; U1151 gene; U1152 gene; U1153 gene; U1154 gene; U1155 gene; U1156 gene; U1157 gene; U1158 gene; U1159 gene; U1160 gene; U1161 gene; U1162 gene; U1163 gene; U1164 gene; U1165 gene; U1166 gene; U1167 gene; U1168 gene; U1169 gene; U1170 gene; U1171 gene; U1172 gene; U1173 gene; U1174 gene; U1175 gene; U1176 gene; U1177 gene; U1178 gene; U1179 gene; U1180 gene; U1181 gene; U1182 gene; U1183 gene; U1184 gene; U1185 gene; U1186 gene; U1187 gene; U1188 gene; U1189 gene; U1190 gene; U1191 gene; U1192 gene; U1193 gene; U1194 gene; U1195 gene; U1196 gene; U1197 gene; U1198 gene; U1199 gene; U1200 gene; U1201 gene; U1202 gene; U1203 gene; U1204 gene; U1205 gene; U1206 gene; U1206.5 gene; U1207 gene; U1208 gene; U1209 gene; U1210 gene; U1211 gene; U1212 gene; U1213 gene; U1214 gene; U1215 gene; U1216 gene; U1217 gene; U1218 gene; U1219 gene; U1220 gene; U1221 gene; U1222 gene; U1223 gene; U1224 gene; U1225 gene; U1226 gene; U1226.5 gene; U1227 gene; U1228 gene; U1229 gene; U1230 gene; U1231 gene; U1232 gene; U1233 gene; U1234 gene; U1235 gene; U1236 gene; U1237 gene; U1238 gene; U1239 gene; U1240 gene; U1241 gene; U1242 gene; U1243 gene; U1244 gene; U1245 gene; U1246 gene; U1247 gene; U1248 gene; U1249 gene; U1250 gene; U1251 gene; U1252 gene; U1253 gene; U1254 gene; U1255 gene; U1256 gene; U1257 gene; U1258 gene; U1259 gene; U1260 gene; U1261 gene; U1262 gene; U1263 gene; U1264 gene; U1265 gene; U1266 gene; U1267 gene; U1268 gene; U1269 gene; U1270 gene; U1271 gene; U1272 gene; U1273 gene; U1274 gene; U1275 gene; U1276 gene; U1277 gene; U1278 gene; U1279 gene; U1280 gene; U1281 gene; U1282 gene; U1283 gene; U1284 gene; U1285 gene; U1286 gene; U1287 gene; U1288 gene; U1289 gene; U1290 gene; U1291 gene; U1292 gene; U1293 gene; U1294 gene; U1295 gene; U1296 gene; U1297 gene; U1298 gene; U1299 gene; U1300 gene; U1301 gene; U1302 gene; U1303 gene; U1304 gene; U1305 gene; U1306 gene; U1307 gene; U1308 gene; U1309 gene; U1310 gene; U1311 gene; U1312 gene; U1313 gene; U1314 gene; U1315 gene; U1316 gene; U1317 gene; U1318 gene; U1319 gene; U1320 gene; U1321 gene; U1322 gene; U1323 gene; U1324 gene; U1325 gene; U1326 gene; U1326.5 gene; U1327 gene; U1328 gene; U1329 gene; U1330 gene; U1331 gene; U1332 gene; U1333 gene; U1334 gene; U1335 gene; U1336 gene; U1337 gene; U1338 gene; U1339 gene; U1340 gene; U1341 gene; U1342 gene; U1343 gene; U1344 gene; U1345 gene; U1346 gene; U1347 gene; U1348 gene; U1349 gene; U1350 gene; U1351 gene; U1352 gene; U1353 gene; U1354 gene; U1355 gene; U1356 gene; U1357 gene; U1358 gene; U1359 gene; U1360 gene; U1361 gene; U1362 gene; U1363 gene; U1364 gene; U1365 gene; U1366 gene; U1367 gene; U1368 gene; U1369 gene; U1370 gene; U1371 gene; U1372 gene; U1373 gene; U1374 gene; U1375 gene; U1376 gene; U1377 gene; U1378 gene; U1379 gene; U1380 gene; U1381 gene; U1382 gene; U1383 gene; U1384 gene; U1385 gene; U1386 gene; U1387 gene; U1388 gene; U1389 gene; U1390 gene; U1391 gene; U1392 gene; U1393 gene; U1394 gene; U1395 gene; U1396 gene; U1397 gene; U1398 gene; U1399 gene; U1400 gene; U1401 gene; U1402 gene; U1403 gene; U1404 gene; U1405 gene; U1406 gene; U1407 gene; U1408 gene; U1409 gene; U1410 gene; U1411 gene; U1412 gene; U1413 gene; U1414 gene; U1415 gene; U1416 gene; U1417 gene; U1418 gene; U1419 gene; U1420 gene; U1421 gene; U1422 gene; U1423 gene; U1424 gene; U1425 gene; U1426 gene; U1426.5 gene; U1427 gene; U1428 gene; U1429 gene; U1430 gene; U1431 gene; U1432 gene; U1433 gene; U1434 gene; U1435 gene; U1436 gene; U1437 gene; U1438 gene; U1439 gene; U1440 gene; U1441 gene; U1442 gene; U1443 gene; U1444 gene; U1445 gene; U1446 gene; U1447 gene; U1448 gene; U1449 gene; U1450 gene; U1451 gene; U1452 gene; U1453 gene; U1454 gene; U1455 gene; U1456 gene; U1457 gene; U1458 gene; U1459 gene; U1460 gene; U1461 gene; U1462 gene; U1463 gene; U1464 gene; U1465 gene; U1466 gene; U1467 gene; U1468 gene; U1469 gene; U1470 gene; U1471 gene; U1472 gene; U1473 gene; U1474 gene; U1475 gene; U1476 gene; U1477 gene; U1478 gene; U1479 gene; U1480 gene; U1481 gene; U1482 gene; U1483 gene; U1484 gene; U1485 gene; U1486 gene; U1487 gene; U1488 gene; U1489 gene; U1490 gene; U1491 gene; U1492 gene; U1493 gene; U1494 gene; U1495 gene; U1496 gene; U1497 gene; U1498 gene; U1499 gene; U1500 gene; U1501 gene; U1502 gene; U1503 gene; U1504 gene; U1505 gene; U1506 gene; U1507 gene; U1508 gene; U1509 gene; U1510 gene; U1511 gene; U1512 gene; U1513 gene; U1514 gene; U1515 gene; U1516 gene; U1517 gene; U1518 gene; U1519 gene; U1520 gene; U1521 gene; U1522 gene; U1523 gene; U1524 gene; U1525 gene; U1526 gene; U1527 gene; U1528 gene; U1529 gene; U1530 gene; U1531 gene; U1532 gene; U1533 gene; U1534 gene; U1535 gene; U1536 gene; U1537 gene; U1538 gene; U1539 gene; U1540 gene; U1541 gene; U1542 gene; U1543 gene; U1544 gene; U1545 gene; U1546 gene; U1547 gene; U1548 gene; U1549 gene; U1550 gene; U1551 gene; U1552 gene; U1553 gene; U1554 gene; U1555 gene; U1556 gene; U1557 gene; U1558 gene; U1559 gene; U1560 gene; U1561 gene; U1562 gene; U1563 gene; U1564 gene; U1565 gene; U1566 gene; U1567 gene; U1568 gene; U1569 gene; U1570 gene; U1571 gene; U1572 gene; U1573 gene; U1574 gene; U1575 gene; U1576 gene; U1577 gene; U1578 gene; U1579 gene; U1580 gene; U1581 gene; U1582 gene; U1583 gene; U1584 gene; U1585 gene; U1586 gene; U1587 gene; U1588 gene; U1589 gene; U1590 gene; U1591 gene; U1592 gene; U1593 gene; U1594 gene; U1595 gene; U1596 gene; U1597 gene; U1598 gene; U1599 gene; U1600 gene; U1601 gene; U1602 gene; U1603 gene; U1604 gene; U1605 gene; U1606 gene; U1607 gene; U1608 gene; U1609 gene; U1610 gene; U1611 gene; U1612 gene; U1613 gene; U1614 gene; U1615 gene; U1616 gene; U1617 gene; U1618 gene; U1619 gene; U1620 gene; U1621 gene; U

FEATURES	source
misc_feature	1..154746 /organism="human herpesvirus 2" /strain="HG52" /db_xref="taxon:10310"
misc_feature	1..254 /note="a sequence"
misc_feature	235 /note="start of TRL (terminal copy of Long Repeat region)"
exon	440..934 /gene="RL1" /number=1 440..126810 /gene="RL1" join(440..934,1089..1379) /gene="RL1" /codon_start=1 /protein_id="CAB06759.1" /db_xref="PID:e30839" /db_xref="PID:g1869821" /db_xref="GI:1869821" /db_xref="SWISS-PROT:P28283"
exon	446..460 /rpt_family="7" 461..547 /rpt_family="6" 943..1070 /rpt_family="5" 1089..1379 /gene="RL1" /number=2 /product="neurovirulence factor"
exon	1738..11743 /gene="RL1" 2303..124947 /gene="RL2" 2303..2377 /gene="RL2" /number=1 join(2303..2377,2785..3462,3644..5368) /gene="RL2" /codon_start=1 /protein_id="CAB06760.1" /db_xref="PID:e304156" /db_xref="PID:g1869822" /db_xref="GI:1869822" /db_xref="SWISS-PROT:P28284"
exon	translation="MEPRPGSSRADGGPEPRPPTGPTOPAPAHAGMLNDKMWLAS SDSEPEVGISDDDLHSDSESGSDTMEFAGLMDAATPRAPRAEROGSTPPAD AOGCGGCPGEEPEAEKGGGDDVCACVCTDEAPLRQCSFPCLHPFCIPCKWIPLR NTCECNTPAAVILIVGTASGSFSTIIVNDFRTVRAEAAVRAGTAVDIITWGNPT
exon	4 (bases 1 to 154746) Barnett,B.C., Dolan,A., Telford,E.A., Davison,A.J. and McGeoch,D.J. A novel herpes simplex virus gene (UL49a) encodes a putative membrane protein with counterparts in other herpesviruses J. Gen. Virol. 73 (Pt 8), 2167-2171 (1992) 5 (bases 1 to 154746) Dolan,A. Direct Submission Submitted (22-FEB-1997) A. Dolan, MRC Virology Unit, Church Street, Glasgow, G11 5JR, UK Location/Qualifiers 1..154746 /organism="human herpesvirus 2" /strain="HG52" /db_xref="taxon:10310" 1..254 /note="a sequence" 235 /note="start of TRL (terminal copy of Long Repeat region)" 440..934 /gene="RL1" /number=1 440..126810 /gene="RL1" join(440..934,1089..1379) /gene="RL1" /codon_start=1 /protein_id="CAB06759.1" /db_xref="PID:e30839" /db_xref="PID:g1869821" /db_xref="GI:1869821" /db_xref="SWISS-PROT:P28283" 446..460 /rpt_family="7" 461..547 /rpt_family="6" 943..1070 /rpt_family="5" 1089..1379 /gene="RL1" /number=2 /product="neurovirulence factor" 1738..11743 /gene="RL1" 2303..124947 /gene="RL2" 2303..2377 /gene="RL2" /number=1 join(2303..2377,2785..3462,3644..5368) /gene="RL2" /codon_start=1 /protein_id="CAB06760.1" /db_xref="PID:e304156" /db_xref="PID:g1869822" /db_xref="GI:1869822" /db_xref="SWISS-PROT:P28284" translation="MEPRPGSSRADGGPEPRPPTGPTOPAPAHAGMLNDKMWLAS SDSEPEVGISDDDLHSDSESGSDTMEFAGLMDAATPRAPRAEROGSTPPAD AOGCGGCPGEEPEAEKGGGDDVCACVCTDEAPLRQCSFPCLHPFCIPCKWIPLR NTCECNTPAAVILIVGTASGSFSTIIVNDFRTVRAEAAVRAGTAVDIITWGNPT

APRS:SLGHTVYRALSPFPMPGDTDEDDIDADVDYVPAPRAAPRRCGGAGATKCT  
SOPATRAPAPGAPSSSSGGAPLNAAGSSGGGPPAAVAAPVPAASLPAPAGAGRAQ  
ARVEEDRAAPAGERTPPAPROPRAAOEPITVSDSGPPSPRAPAGPLSEFSSSSAAQV  
SSPBGGGGIPGSSGGAAPRAAIVAPRVSPSPRAAAPVVASADAAGAPAPVPAVAKA  
RSPRRGQAOADPDNOAGIAGRATPDAGSGCGPAGPGVPGCTNTPGAPAAVAKA  
AARPKRRGSDSGPAASSSSAASNAAPRPLAPQVGAKRAPRAPRPSDSDBDRHGTPL  
AASAGGAAPSPASPSQAALVAASSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS  
SSSSAGGAAPVASAGAEERETSLGPPAAPRBPCKARTRHAEGGPEGGADPPAP  
GTRILPLAGSSSVVALAPVYKTVTGCLPLVDMETGHIAGAYVLVDQETNVADLIR  
AAAPASRRITLPEHARNCVPRPDYPTPASEWMSLWMTPGNNLFPDGLVGLDFH  
GIRSHHPSREGCAPACGADAPAGHGE"  
2785..3462  
/gene="RL2"  
/number=2  
3468..3607  
/rpt\_family="4"  
3644..5368  
/gene="RL2"  
/number=3  
/function="modulator of cell state and gene expression"  
4655..4771  
/rpt\_family="3"  
complement(4811)  
/gene="RL2"  
/note="proposed LAT splice acceptor site"  
5618..5623  
/gene="RL2"  
complement(7025)  
/gene="RL2"  
/note="proposed LAT splice donor site (5' end of stable  
LAT species)"  
complement(7732)  
/gene="RL2"  
/note="primary transcription initiation site for latency  
associated transcripts (LAT's)"  
9047..9193  
/rpt\_family="2"  
9276..9297  
/rpt\_family="1"  
9298  
/gene="RL2"  
/note="start of UL (long unique region)"  
9427..10101  
/gene="UL1"  
9427..10101  
/codon\_start=1  
/product="viral glycoprotein L"  
/protein\_id="CAB06761.1"  
/db\_xref="PID:e304262"  
/db\_xref="PID:g1869823"  
/db\_xref="GI:1869823"  
/db\_xref="Swiss-Prot:P28278"  
/translation="MGVCLPGLVVMGAMGGSQATEYLRVSIAKEVGDILRVPCC  
MHTPQDSVSMRYEAPSVIDYARIDGIFLRKYCPGLDTFLMDRHAORAYLVNPFSTAG  
FLEIDSHVPEADQVETTRTALYKKEITDALGSRKQAVSHAPVAGCVNDYSTTRC  
VGRRLRPANNTSTMEPPVSSDDASQSKRLAQPVYALSNAPPRVSPTRGRKH  
TRLRN"  
10211..10978  
/gene="UL2"  
10211..10978  
/gene="UL2"  
/codon\_start=1  
/product="viral DNA glycosylase"  
/protein\_id="CAB06762.1"  
/db\_xref="PID:e304157"  
/db\_xref="PID:g1869824"  
/db\_xref="GI:1869824"  
/db\_xref="Swiss-Prot:P28275"  
/translation="MFSASTPEOPLGSDATPEPTSVPLDMAFRAFLIDDAWR  
PLLEELNPLTALRLAIEYDRRCOTIEVLPRDEVFSWTRCTPDDVRAVYIIGDPPH  
HPGQAHGAFSEVRADVPVPSLRNVLAIVKNCYEDAMSGRGCELEKARDAVGLLYNTT  
LTVRGAASASISKIGMDRFVGGVQRLAARPGVFLMGAHQONAIRPVRPHQYVYK





cds /number=1  
join(2303..2377,2785..3462,3644..5368)  
/gene="RL2"  
/codon\_start=1  
/protein\_id="CAB06760.1"  
/db\_xref="PID:e304156"  
/db\_xref="PID:g1869822"  
/db\_xref="GI:1869822"  
/db\_xref="SWISS-PROT:P28278"  
/translation="MEPRGTSRADGPPRRPPTGTPAPAHMGLINDQWLAS  
SDEETEVGIIDDLDHROSTSEAGSTDEIAPGLMDAATPARPAERKQSPAD  
ADSCCGGPGVEEABAGGGGDCVACDTIAPLCOSPPCLHPCIDMKMILR  
NTPLCNTPAYALIVGTASGSESTPIVNDPRTVEAEAAVAGTAVDEITNGNRT  
APRLSLGHTVRLALPTPPMPTDDEDDLDADVDVPPAPRRAPRRGGGAGATGCT  
SQPATRRPAPRRSSSGGALIRAGVSGSGGCAVAAYVRVSLPPRAGGGAQ  
ARRVGDAAAEGRTPPARQAPRAQEPPIYISDPSPPSPRPGPPLSFVSSSAQV  
SSQGGGGGLQSSGGAARPAARAAVPRVSPRAAAPVVSADAAAGPAPVADAH  
RAVSRMTAQDTQOASLRAGATDAGSGGAGGPGVPGTNTPGAPAAAGAA  
AAPRRRGSDSGPAASSASASAPRSPPLAPGAGKRAAPRRAPRRDGGDHGPL  
APASAGAAPPSPSSOAAVAAASSSASSSASSSASSSASSSASSSASSSA  
SSAGAGGSVASASGAGERETSLPRAAPRPGPKCAKTHAGGPPRANRDPAP  
GLTRILPIAGVSVVALAPYVNTVTGDCPLVDMETGHIGATVVLVDGTGNADLLR  
AAPAWSRRLLEHARNVCYRPDPYPTPPASEMNSIMTPEVGNMLFDQGLTVALDFH  
GLSRHPMSREOGAPAPAGDAPAGHGE"  
2785..3462  
/gene="RL2"  
/number=2  
3468..3607  
/rpt\_family="4"  
3644..5368  
/gene="RL2"  
/number=3  
/function="modulator of cell state and gene expression"  
4655..4771  
/rpt\_family="3"  
complement(4811)  
/gene="RL2"  
/note="proposed LAT splice acceptor site"  
5618..5623  
/gene="RL2"  
complement(7025)  
/gene="RL2"  
/note="proposed LAT splice donor site (5' end of stable  
LAT species)"  
complement(7732)  
/gene="RL2"  
/note="primary transcription initiation site for latency  
associated transcripts (LAT's)"  
9047..9193  
/rpt\_family="2"  
9276..9297  
/rpt\_family="1"  
9298  
/gene="RL2"  
/note="start of UL (Long Unique region)"  
9427..10101  
/gene="UL1"  
9427..10101  
/gene="UL1"  
/codon\_start=1  
/product="viral glycoprotein L"  
/protein\_id="CAB06761.1"  
/db\_xref="PID:e304262"  
/db\_xref="PID:g1869823"  
/db\_xref="GI:1869823"  
/db\_xref="SWISS-PROT:P28278"  
/translation="KGFVCLFGLVYVANGAMGSGATEVYVLSVTAKEVGLTRVPC  
MRTPADVSWRTAPSVIDARIDGILFIRHCGGLDTPLMDRHAQRAYLVNPLFLAG  
FLDLDSHSVFPADOTETTRALYKEIRDALGSRKQAVSHAPVACGVNEDYSRTRRR  
VGRDLRPANTSTWEPVSSDDEASSOSKPLATOPVLAISNAPRRVSPTRGRRRH  
TRLRRN"  
10211..10978  
/gene="UL2"

cds 10211..10978  
/gene="UL2"  
/codon\_start=1  
/product="uracil-DNA glycosylase"  
/protein\_id="CAB06762.1"  
/db\_xref="PID:e304157"  
/db\_xref="PID:g1869824"  
/db\_xref="GI:1869824"  
/db\_xref="SWISS-PROT:P28275"  
/translation="MFSASTPEOPPLGSDATPLPVSPLDMAAFRRAPLIDAMR  
PLLEPLANPLTARILAEYDRRCQTEEVYPPREDVFSWTRYPDPDVRVILIGOPYH  
HGOAHLAFSVRAVPPSLRNLYLAANKCYPARMSRGCLERKADGVLNNT  
LVKRGAAASHRGKIDRWGVGVORLARRGGLVFMGMGAIONAIRDPQHYLK  
FHSRPSLSVPGTCOHFLAANRYLETDRIMEIDMSV"  
11019..11024  
/gene="RL2"  
11033..11734  
/gene="UL3"  
11033..11734  
CDS  
Query Match 16.1%; Score 63; DB 17; Length 154746;  
Best Local Similarity 48.7%; Pred. No. 0.083;  
Matches 171; Conservative 0; Mismatches 180; Indels 0; Gaps 0;  
QY 37 GCCCTGCGCTGGCCAGAGCGCGGAGCCGACAGCCCGCTCGGTGAGACTGGGGGTGCA 96  
Db 148472 GCCGGGGGAGCGCGCGCGGGGAGCGCGCGGGGAGCGCGCGGGGAGCGCGGGG 148413  
QY 97 GGTGCCGAGACCTACCCAGCGCGGAGTACCTCCCGCTCACACCTCGGGCTGCAGTTG 156  
Db 148412 GACGGGCGCGGGGAGCGCGCGGGGAGCGCGCGGGGAGCGCGGGGAGCGCGGGG 148353  
QY 157 CCTGTGTCGCCCGCGAGACGCTGCGCCCGGCTGAGAGGATGCGGGGAGCGGGAGCG 216  
Db 148352 GCCGGGGGCGCGGGGCGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCG 148293  
QY 217 GGGGGCGGGCTCTGCTACGTGAGAGCGCGCGGGGAGCGGGGAGCGGGGAGCGGGG 276  
Db 148292 GGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGG 148233  
QY 277 CGGCTCCTTAAAGCGCGGAGCGCGGAGCGCGGAGTGTCTCTGTGGCCGCGGAGCGG 336  
Db 148232 GACGGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGG 148173  
QY 337 CCGGGGAGCGCGGAGACGTGAGCGCGCGCTCTGCTGAGCTCCGGGAGC 387  
Db 148172 CGGGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCG 148122  
RESULT 12  
CH19R30879/c  
LOCUS CH19R30879 41936 bp DNA PRI 01-APR-1997  
DEFINITION Homo sapiens DNA from chromosome 19- cosmid R30879 containing USR2,  
genomic sequence.  
ACCESSION AD000684  
NID AD000684  
VERSION 91905917  
KEYWORDS AD000684.1 GI:1905917  
SOURCE chromosome 19; transcription factor.  
ORGANISM human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 41936)  
AUTHORS Lamerdin,J.E.  
TITLE Direct Submision  
JOURNAL Submitted (10-DEC-1996) J.E. Lamerdin, Human Genome Center,  
Lawrence Livermore National Laboratory, 7000 East Ave, Livermore,  
CA, USA, 94551 janefecg.1lnl.gov owetornak.1lnl.gov  
COMMENT GSDB:S:1010600.  
Human Genome Center  
Biology and Biotechnology Research Program  
Lawrence Livermore National Laboratory  
7000 East Avenue

Livermore, CA 94550 USA  
constructed at LNL from flow-sorted chromosomes  
from hybrid 5HL2-B, which carries chromosome 19 as its only human  
chromosome.

## FEATURES

Location/Qualifiers  
1. 41936  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/cell\_line="5HL2-B"  
/cell\_type="fibroblast"  
/clone="R30879"  
/clone\_id="L119NC03 R chromosome 19-specific cosmid  
library"  
repeat\_region  
complement(1..34)  
/note="repeat match = HSAL15306; putative"  
/rpt\_family="Alu"  
repeat\_region  
complement(811..1099)  
/note="repeat match = HSAL00208; putative"  
/rpt\_family="Alu"  
repeat\_region  
1335..1479  
/note="repeat match = HSAL05512; putative"  
/rpt\_family="Alu"  
repeat\_region  
1367..1305  
/note="repeat match = HSAL02043; putative"  
/rpt\_family="Alu"  
repeat\_region  
complement(3093..3356)  
/note="repeat match = HSAL02254; putative"  
/rpt\_family="Alu"  
repeat\_region  
4030..4168  
/note="repeat match = HSAL03448; putative"  
/rpt\_family="Alu"  
repeat\_region  
4031..4288  
/note="repeat match = HSAL01019; putative"  
/rpt\_family="Alu"  
repeat\_region  
4435..4533  
/note="repeat match = HSAL02522; putative"  
/rpt\_family="Alu"  
repeat\_region  
4449..4565  
/note="repeat match = HSAL01715; putative"  
/rpt\_family="Alu"  
repeat\_region  
4926..5218  
/note="repeat match = HSAL02523; putative"  
/rpt\_family="Alu"  
repeat\_region  
complement(5804..6091)  
/note="repeat match = HSAL04878; putative"  
/rpt\_family="Alu"  
repeat\_region  
complement(6051..6102)  
/note="repeat match = HSAL04280; putative"  
/rpt\_family="Alu"  
repeat\_region  
7131..7434  
/note="repeat match = HSAL02525; putative"  
/rpt\_family="Alu"  
repeat\_region  
7137..7589  
/note="repeat match = HSAL00829; putative"  
/rpt\_family="Alu"  
repeat\_region  
7911..16824  
/gene="LISCH7"  
join(<7911..8030,10891..10947,11512..11658,15325..15498,  
15803..15933,16075..16700,16790..16824)  
/gene="LISCH7"  
/function="human homolog of Mus musculus BGCBA Lisch7  
mRNA"  
/note="putative"  
/codon\_start=2  
/product="liver-specific BHLH-zip transcription factor"  
/protein\_id="AAB51178.1"  
/db\_xref="PID:q1905918"  
/db\_xref="GI:1905918"  
/translation="ADLTPTQATGSGVYCSVNAQDIQNNAYAEIIVIGRSG  
VAELIFGAGPIEGWLVVAVCIAPLILLLGICWCQCPCPTCCCYVGPCPCPC  
CCPAPYAAKAKATSGVSIYAPSTYAPLIPKPPAPMIPGAPYNGYPCGYPDCV  
DRSSGVSQYRIQASQDDSKMVLVYMERELANFDSRPGSPGRVGRMSVETSLH

repeat\_region  
8303..8386  
/note="repeat match = HSAL05819; putative"  
/rpt\_family="Alu"  
repeat\_region  
8325..8609  
/note="repeat match = Alu; putative"  
/rpt\_family="Alu"  
repeat\_region  
complement(9237..9546)  
/note="repeat match = HSAL05084; putative"  
/rpt\_family="Alu"  
repeat\_region  
complement(9513..9555)  
/note="repeat match = HSAL00054; putative"  
/rpt\_family="Alu"  
repeat\_region  
9599..9759  
/note="repeat match = HSAL17230; putative"  
/rpt\_family="Alu"  
repeat\_region  
9639..9813  
/note="repeat match = HSAL06130; putative"  
/rpt\_family="Alu"  
repeat\_region  
10359..10498  
/note="repeat match = HSAL06554; putative"  
/rpt\_family="Alu"  
repeat\_region  
11510..11657  
/gene="LISCH7"  
/note="similarity: gb|H421281|H42128.y061b12.r1 Homo  
sapiens cDNA clone 182399 5'; putative"  
/note="repeat match = HSAL01057; putative"  
/rpt\_family="Alu"  
repeat\_region  
11981..12253  
/note="repeat match = Alu; putative"  
/rpt\_family="Alu"  
repeat\_region  
12168..12279  
/note="repeat match = HSAL03934; putative"  
/rpt\_family="Alu"  
repeat\_region  
complement(12334..12610)  
/note="repeat match = HSAL00279; putative"  
/rpt\_family="Alu"  
repeat\_region  
complement(12391..12982)  
/note="repeat match = HSAL02441; putative"  
/rpt\_family="Alu"  
repeat\_region  
complement(12645..13088)  
/note="repeat match = HSAL02504; putative"  
/rpt\_family="Alu"  
repeat\_region  
16521..16586  
/gene="LISCH7"  
/note="similarity: gb|R39945|R39945.yf52e09.s1 Homo  
sapiens cDNA clone 25701 3'; similar to . . .; putative"  
16573..16699  
/gene="LISCH7"  
/note="similarity: gb|R36881|R36881.yf52a08.s1 Homo  
sapiens cDNA clone 25687 3'; . . .; putative"  
16577..16699  
/gene="LISCH7"  
/note="similarity: gb|R39945|R39945.yf52e09.s1 Homo  
sapiens cDNA clone 25701 3'; similar to . . .; putative"  
16594..16650  
/gene="LISCH7"  
/note="similarity: gb|H06137|H06137.y177c10.s1 Homo  
sapiens cDNA clone 44249 3'; . . .; putative"  
16646..16699  
/gene="LISCH7"  
/note="similarity: gb|H06137|H06137.y177c10.s1 Homo  
sapiens cDNA clone 44249 3'; . . .; putative"  
16785..16867  
/note="similarity: gb|H06137|H06137.y177c10.s1 Homo  
sapiens cDNA clone 44249 3'; . . .; similarity:  
gb|R39945|R39945.yf52e09.s1 Homo sapiens cDNA clone 25701  
3'; similar to . . .; similarity: gb|R36881|R36881.  
yf52a08.s1 Homo sapiens cDNA clone 25687 3'; . . .;

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

misc\_feature

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

```

misc_feature      putative"
16852..16907
/note="similarity: gb|H06137|H06137; Y177C10.s1 Homo
sapiens cDNA clone 44249 3' . . . ; putative"
16852..16877
/note="similarity: gb|R39945|R39945; yf52e09.s1 Homo
sapiens cDNA clone 25701 3' similar to . . . ; similarity:
gb|R36881|R36881; yf52a08.s1 Homo sapiens cDNA clone 25687
3' . . . ; putative"
16883..16918
/note="similarity: gb|R39945|R39945; yf52e09.s1 Homo
sapiens cDNA clone 25701 3' similar to . . . ; putative"
16896..16929
/note="similarity: gb|R36881|R36881; yf52a08.s1 Homo
sapiens cDNA clone 25687 3' . . . ; putative"
18067..28221
/gene="USF2"
join(18067..18128,18411..18457,18546..18664,18768..18968,
19412..19562,19683..19770,20048..20106,27663..27757,
27911..28039,28132..28221)
/gene="USF2"
/function="ubiquitous transcription factor"
/note="putative"
/codon_start=1
/product="upstream stimulatory factor 2"
/protein_id="AAB51179.1"
/db_xref="PIR:G1805919"
/db_xref="GI:1905919"
/translaton="MDMLDPLDPASATATAAASHDKGPEAEVGLQEGDGPAGAE
EOTAVATSYOQAAGFDHNIYOFRETNGGOVTVRVVOTDGLDGGDTAGAVSV
STAFAPGGOAVTVGVDAAGORGPAAASVPGPAAPPLAVIONPFGNSGSPAARA
VSGEARFATFPASSVCDTTAVSVOTDQSLQAGGQFYVMATPDVLTGTORTIART
VSGEARFATFPASSVCDTTAVSVOTDQSLQAGGQFYVMATPDVLTGTORTIART

Query Match      15.7%; Score 61.6; DB 10; Length 41936;
Best Local Similarity 53.1%; Pred. No. 0.2;
Matches 155; Conservative 0; Mismatches 134; Indels 3; Gaps 1;

QY 55 GCGCGAGAGCCGAGAGCCGCTCGGTGAGACTGGGGGTGAGTCCCGAGACCTTACC 114
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18186 GCGCGAGAGCCGAGAGCCGCTCGGTGAGACTGGGGGTGAGTCCCGAGACCTTACC 18127

QY 115 AGCGCGGAGATACCTCCGCTCACACCTCGGCGTCACTCCGCTGAGCGCGAGAGA 174
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18126 AGCGCGGAGATACCTCCGCTCACACCTCGGCGTCACTCCGCTGAGCGCGAGAGA 18070

QY 175 CGCTGCGCGCGGCTGAGAGATGCGGGGCGGGAAGCGGGGCGGGGCTGCTCAC 234
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18069 CATGGGGGCGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGAG 18010

/ 235 GTGAGAGAGCGCGCGGGGCGGGGCGGGGCGGGGCGGGGCTCTTAAGCGCG 294
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18009 GAGAGGAGGAGGAGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCTCACGCGCG 17950

QY 295 CGAGCCGAGCGGAGGTCTGTGTGCGCCAGGCGGAGGCGCGCGAGACA 346
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17949 GCAGAGGAGGAGCGAGGATACCGGAGCGCGCTCGCGCTGTATCAGGAGACA 17898

RESULT 13
HSUSF2/c          HSUSF2      14440 bp      DNA          PRI          26-JUN-1997
LOCUS             H.sapiens USF2 gene.
DEFINITION        Y07661
ACCESSION         Y07661
NID               G1806093
VERSION           Y07661.1 GI:1806093
KEYWORDS          USF2 gene.
SOURCE            human.
ORGANISM          Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
REFERENCE         1 (Bases 1 to 14440)
AUTHORS           Groenen,P.M., Garcia,E., Debeer,P., Devriendt,K., Fyng,J.P. and
Van de Ven,W.J.

```

```

TITLE             Structure, sequence, and chromosome 19 localization of human USF2
                  and its rearrangement in a patient with multicystic renal dysplasia
JOURNAL           Genomics 38 (2), 141-148 (1996)
MEDLINE           97127588
REFERENCE         2 (Bases 1 to 14440)
AUTHORS           Groenen,P.
TITLE             Submitted (27-AUG-1996) P. Groenen, Center for Human Genetics,
                  Laboratory for Molecular Oncology, Herestraat 49, 3000 Leuven,
                  BELGIUM
JOURNAL           Related sequence: X90823-X90826.
COMMENT           location/Qualifiers
FEATURES          1..14440
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="19"
                  /haplotype="2n"
                  /map="q13.1"
                  /clone_1lb="L1NT.7058"
                  /dev_stage="adult"
                  join(3688..3749,4031..4077,4166..4284,4388..4588,
5032..5182,5303..5390,5668..5726,13284..13378,
13532..13660,13753..14380)
                  /gene="USF2"
                  3688..14380
                  /gene="USF2"
                  3688..3749
                  /number=1
                  join(3688..3749,4031..4077,4166..4284,4388..4588,
5032..5182,5303..5390,5668..5726,13284..13378,
13532..13660,13753..13842)
                  /gene="USF2"
                  /codon_start=1
                  /protein_id="CA68942.1"
                  /db_xref="PIR:G291919"
                  /db_xref="PIR:G1806094"
                  /db_xref="GI:1806094"
                  /db_xref="SWISS-PROT:Q15853"
                  /translaton="MDMLDPLDPASATATAAASHDKGPEAEVGLQEGDGPAGAE
EOTAVATSYOQAAGFDHNIYOFRETNGGOVTVRVVOTDGLDGGDTAGAVSV
STAFAPGGOAVTVGVDAAGORGPAAASVPGPAAPPLAVIONPFGNSGSPAARA
VSGEARFATFPASSVCDTTAVSVOTDQSLQAGGQFYVMATPDVLTGTORTIART
HPSFPRIDTFRTPRDRRRAROHNEVERERLDINNIIVOLSKTIPPCNADNSTGASK
GGILSKACDIYRELRTQNMQETFEARERLDQMLELROQIEELKERNALLRAQLQQ
HNLEWEGEGTRQ"
                  3750..4030
                  /gene="USF2"
                  /number=1
                  4031..4077
                  /gene="USF2"
                  /number=2
                  4078..4165
                  /gene="USF2"
                  /number=2
                  4166..4284
                  /gene="USF2"
                  /number=3
                  4285..4387
                  /gene="USF2"
                  /number=3
                  4388..4588
                  /gene="USF2"
                  /number=4
                  4589..5031
                  /gene="USF2"
                  /number=4
                  5032..5182
                  /gene="USF2"
                  /number=5
                  5183..5302
                  /gene="USF2"
                  /number=5

```

```

exon      5303..5390
           /gene="USF2"
           /number=6
Intron    5391..5667
           /gene="USF2"
           /number=6
exon      5668..5726
           /gene="USF2"
           /number=7
Intron    5727..13283
           /gene="USF2"
           /number=7
exon      13284..13378
           /gene="USF2"
           /number=8
Intron    13379..13531
           /gene="USF2"
           /number=8
exon      13532..13660
           /gene="USF2"
           /number=9
Intron    13661..13752
           /gene="USF2"
           /number=9
exon      13753..14380
           /gene="USF2"
           /number=10
BASE COUNT 2766 a 4210 c 4353 g 3110 t 1 others
ORIGIN

```

```

Query Match      15.7%   Score 61.6; DB 10; Length 14440;
Best Local Similarity 53.1%; Pred. No. 0.29;
Matches 155; Conservative 0; Mismatches 134; Indels 3; Gaps 1;

```

```

QY  55 GCGCGGAGGCGCCAGAGCCCGCTCGGTGAGAGACTGGGGGTGAGAGTCCCGGAGCGTACC 114
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  3807 GCGCGGAGGCGCCAGAGCCCGCTCGGTGAGAGACTGGGGGTGAGAGTCCCGGAGCGTACC 3748
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  115 AGCGCGGAGAGTACTCCCGCTCACACCTCGGCTGACATTCTGTGGGCGCGCGAGAGA 174
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  3747 GCGCGGAGAGAGCGGTGCGCGAGGAGCGAGCGGATCCAGACCC---GGGTCCAGATGTC 3691
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  175 CGCTGCGCCCGGCTGAGAGGATGCGGGGCGGAGCGGGGCGGGGCGGCTGTCAC 234
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  3690 CATGGGAGGAGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGAGGAGGAGGAG 3631
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  235 CTGGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTTAAGCGCG 294
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  3630 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3571
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  295 CGAGCGGAGCGCGGAGGTCTGTGTGCGCGAGGCGAGCGCGCGCGCGAGCA 346
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  3570 GCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCA 3519
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 14
LOCUS      SCE94
DEFINITION Streptomyces coelicolor cosmid E94.
ACCESSION AL049628
VERSION    94585581
KEYWORDS   AT049628.1 GI:4585581
SOURCE     Streptomyces coelicolor.
ORGANISM   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Streptomyces coelicolor
            Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.

```

REFERENCE  
AUTHORS Redenbach, K., Kleser, H. M., Denapate, D., Eichner, A., Collum, J., Kinashi, H. and Hopwood, D. A.  
TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)  
MEDLINE 9700351  
REFERENCE 2 (bases 1 to 38532)  
AUTHORS Oliver, K. and Harris, D.  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 38532)  
AUTHORS Bentley, S. D., Parkhill, J., Barrell, B. G. and Rajandream, M. A.  
TITLE Direct Submission  
JOURNAL Submitted (12-APR-1999) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT  
Notes: Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC.  
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
(URL: <http://www.sanger.ac.uk/Projects/S-coelicolor/>)  
CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).  
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.  
The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.  
Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.  
Cosmid E94 lies between E7 and E126 on the AseI-E genomic restriction fragment.

FEATURES  
SOURCE location/Qualifiers  
1..38532  
/organism="Streptomyces coelicolor"  
/strain="A3(2)"  
/db\_xref="taxon:1902"  
/clone="cosmid E94"  
complement(1..1512)  
/gene="SCE94.01c"  
1..105  
/gene="SCE94.01c"  
/note="Nominal overlap with Streptomyces coelicolor E7."  
complement(<1..1512)  
/gene="SCE94.01c"  
/note="SCE94.01c, partial CDS unknown, alanine-rich protein, len: >503aa"  
/codon\_start=11  
/label="table-11"  
/product="alanine-rich hypothetical protein"  
/protein\_id="CAB40850.1"  
/db\_xref="PID:e1424928"  
/db\_xref="PID:94585582"

```

/db_xref="GI:4585582"
/translation="MRSSPRTTGTGAHRAHREARDRTAARALAPPARPARYA
VDAYIDGLETVCISLCHDADATATAGVLAIAERGRGRPADRBSMLYALAKMA
CIRKAEKOKRPSHTAHAGRPAPVTROTQTPIATPTINDETHORHRESEALAMP
EAGTTPQREKRELELAHRLHAPHEVAALVGLDIAATRELLASACEVEERFALAVV
ETGACAGARLRLGNRYLSTALRELVHVDCCRCRTIARAIPGRMPGTSVPAAE
LPVLEAPRTALVALAHSTARSAPREFRPFMDPRARARDLRARAATTVVA
TVVAAPVLAIMAYRGTPVEEGEGRSASAPADPDADGDSAGAAIYENAGNAST
TPGTFGEKGGADYSEVYVYSGASGRASAPADPDADGDSAGAAIYENAGNAST
ASVGASMLTSSGSGTLNRGEAVTVRVYDHLREPSGGRSARVAVSPAGAVAT"
2126..3535
/gene="SCE94.02"
2126..3535
/gene="SCE94.02"
2126..3535
/note="SCE94.02, possible DNA repair protein, len: 469aa;
similar to many eg. SW:RADA_ECOLI Rada, DNA repair protein
(mechanism unknown) from Escherichia coli (460 aa) fasta
scores: opt: 1218, z-score: 1309.3, E(): 0, (41.4%
identity in 452 aa overlap). Contains Prosite match to
PS00017 ATP/GTP-binding site motif A (P-loop)."
```

```

/note="SCE94.04c, unknown, len: 126aa"
/codon_start=1
/translation="MRSSPRTTGTGAHRAHREARDRTAARALAPPARPARYA
VDAYIDGLETVCISLCHDADATATAGVLAIAERGRGRPADRBSMLYALAKMA
CIRKAEKOKRPSHTAHAGRPAPVTROTQTPIATPTINDETHORHRESEALAMP
EAGTTPQREKRELELAHRLHAPHEVAALVGLDIAATRELLASACEVEERFALAVV
ETGACAGARLRLGNRYLSTALRELVHVDCCRCRTIARAIPGRMPGTSVPAAE
LPVLEAPRTALVALAHSTARSAPREFRPFMDPRARARDLRARAATTVVA
TVVAAPVLAIMAYRGTPVEEGEGRSASAPADPDADGDSAGAAIYENAGNAST
TPGTFGEKGGADYSEVYVYSGASGRASAPADPDADGDSAGAAIYENAGNAST
ASVGASMLTSSGSGTLNRGEAVTVRVYDHLREPSGGRSARVAVSPAGAVAT"
2126..3535
/gene="SCE94.02"
2126..3535
/gene="SCE94.02"
2126..3535
/note="SCE94.02, possible DNA repair protein, len: 469aa;
similar to many eg. SW:RADA_ECOLI Rada, DNA repair protein
(mechanism unknown) from Escherichia coli (460 aa) fasta
scores: opt: 1218, z-score: 1309.3, E(): 0, (41.4%
identity in 452 aa overlap). Contains Prosite match to
PS00017 ATP/GTP-binding site motif A (P-loop)."
```

subunit.  
ACCESSION 212168  
ID 9311336  
VERSION 212168.1 GI:311336  
KEYWORDS GTP binding protein; GTP binding protein alpha subunit.  
SOURCE dog.  
ORGANISM Canis familiaris  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Carnivora; Fissipedia; Canidae; Canis.  
REFERENCE 1 (bases 1 to 1927)  
AUTHORS Ishikawa, Y. and Homcy, C.J.  
TITLE CDNA sequence for alpha subunit of stimulatory guanine nucleotide  
binding protein from canine heart  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1927)  
AUTHORS Ishikawa, Y.  
TITLE Direct Submission  
JOURNAL Submitted (01-JUN-1992) Yoshihiro Ishikawa, Department of  
Pharmacology, Columbia University, 630 W, 168th St., New York, NY,  
10032, USA

FEATURES  
source Location/Qualifiers  
1..1927  
/organism="Canis familiaris"  
/db\_xref="taxon:9615"  
/tissue\_type="Ventricular heart muscle"  
/clone="G1one 7M"  
354..1538  
/codon\_start=1  
/product="stimulatory GTP binding protein"  
/protein\_id="CAA78161.1"  
/db\_xref="pid:311337"  
/db\_xref="gi:311337"  
/db\_xref="swiss-prot:P04895"  
/translation="MGCLGNSKTEDQNEKAREANKRIEKOLOKQVYRATHRL  
LLGAGESKSTIVKQRIILHVNGEGEEDPQARNSDGEKATKQDIKNINKEA  
IETIVAAASNLVPELANPENOPEVDYILSYNVPDPPPEFEHAKALMEDEGVR  
ACERSNEQLIDCAQYFLDKDYKADYPSDDLLRCRYLTGIFETKQYDKVN  
FMFEPVGGRRDRRRKIOCFNDVYALITFVASSSTINMYIREDNQTRIQEALNLFKSI  
WNNRWLRTISVILFNKQDLAEKVLAKSKIEDYFPEFARTTPEDATPEGDPYV  
TRAKFIIDFELRISTASGDGRHYCPHFTCAVDTEINIRVFNDCRDIIOHMLROYE  
LL"

BASE COUNT 493 a 576 c 500 g 358 t  
ORIGIN

Query Match 15.6%; Score 61; DB 3; Length 1927;  
Best Local Similarity 52.1%; Pred. No. 0.69;  
Matches 159; Conservative 0; Mismatches 145; Indels 1; Gaps 1;

QY 31 TGGGCAAGCCCTGGCTGGCCAGAGCGCGAGCGCCGCTGGTGAGACTGGG 90  
DB 319 TGGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 260  
QY 91 GGTGAGAGTGGCCGAGCGTACCCAGC-GCCGGAGTACCTCCCGCTCACACTGGGCT 149  
DB 259 GGGCGGGAGGGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 200  
QY 150 GCAATTCCCTGGGTGGCCGCGAGACGCTGGCCCGGGCTGAGAGGATGGCGGGCGGGA 209  
DB 199 GAGAGCGCGCGCGCGCGAGTGGCGGGCGCGCGCGCGCGCGCGCGCGG 140  
QY 210 CGGGGCGCGGGGCGGGGCTGCTACGTGAGAGAGCGCGCGGGGCGGGGCGGGGCG 269  
DB 139 CCGAGGGGAGCGCGGAGCTCCCGCCCTCGAGCCGAGGCGGAGGGGCTGATGGCCGC 80  
QY 270 GCGCGCCGCGGCTCTTAAGAGCGCGGAGCGCGAGCGAGGAGTGCCTGTGGCGCGAG 329  
DB 79 CGCGGGGCGGAGCGGAGCCAGAGCAGAGACTGGGGGAGACTGCTGCCCGCGCGCGC 20  
QY 330 CGCAG 334  
DB 19 CGCGG 15

Search completed: September 17, 1999, 22:05:04  
Job time: 16487 sec

Result No.	Score	Match	Length	DB	ID	Description
1	72	18.4	1876	1	T48239	Neuronal nicotinyl
2	72	18.4	1876	1	V12197	Human neuronal ni
3	69.2	17.7	3861	1	Q51543	Lipopolysaccharid
4	64	16.3	3198	1	X02974	Human IL-1ra BAC
5	63.6	16.2	114955	1	X53491	Human adenosine A
6	60.6	15.5	1281	1	Q23298	HSV-1 (MGH-10) IC
7	57.8	14.7	11820	1	V18130	Human chromosome
8	57.8	14.7	114955	1	X53491	Human adenosine
9	56.4	14.4	12001	1	Q76213	HSV L/ST region
10	56	14.3	1120	1	V60339	CDNA sequence of
11	52.4	13.4	12001	1	Q76213	HSV L/ST region
12	51.8	13.2	4403	1	Q47927	HSV L/ST region
13	51.8	13.2	2218	1	Q47929	Paired basic amino
14	51.8	13.2	4257	1	V10362	Paired basic amino
15	51.8	13.2	4257	1	V68520	Infected cell prod
16	51.2	13.1	5228	1	V81384	The nucleotide seq
17	50.4	12.9	7193	1	V50431	Nucleic acid encod
18	50.2	12.8	1335	1	Q23295	Streptomyces clav
19	49.6	12.7	1303	1	Q23297	HSV-1 (F) ICP3.4.5
20	49.6	12.7	1303	1	V44430	HSV-1 (CVG-2) ICP
21	49.6	12.7	407	1	V64530	Mycobacterium tub
22	49.2	12.6	407	1	V64530	M. tuberculosis in
23	49	12.5	701	1	Q76209	HSV L/ST ORF1. Her
24	49	12.5	4405	1	Q25444	PACE composite seq
25	49	12.5	4405	1	T05556	Human PACE coding
26	49	12.5	223	1	T66412	Wild type SNF 5' U
27	48.4	12.3	7011	1	V20464	Human T-cell leuk
28	48.4	12.3	801	1	V73801	Human L-myc oncoge
29	48	12.2	4020	1	T91361	KSHV LTR terminal
30	48	12.2	117213	1	V62176	Orf virus genomic
31	47.8	12.2	1032	1	Q44281	HSV-2 strain SB5 C
32	47.6	12.1	6225	1	X55273	Pseudomonas cepac
33	47.4	12.1	1610	1	Q57657	Human enzyme-rela
34	47.4	12.1	16054	1	T27547	Allele D4.7 of the
35	47.4	12.1	4425	1	X32641	Recombinant human
36	47.2	12.0	117213	1	V62176	Human choroidi
37	46.8	11.9	4468	1	T35520	HSV-2 strain SB5 C
38	46.8	11.9	339	1	V44425	Human thrombopoiet
39	46.8	11.9	339	1	V64534	Mycobacterium tub
40	46.2	11.8	4356	1	Q37543	Human tuberculosis im
41	46.2	11.8	4356	1	Q95540	Cardiac adenylyl c
42	46	11.7	1755	1	Q11243	Cardiac adenylyl c
43	46	11.7	189416	1	V08416	PRV 28K gene. Mutat
						HCW strain AD169

eryA region of S.  
Human dopamine D4

## RESULT 1

ID T48239 standard; DNA; 1876 BP

DT 09-APR-1997 (first entry)

Neuronal nicotinic acetylcholine receptor; nAChR; neurotransmitter,

EH	OS	Key	Location/Owner
		hollo saprens.	

33

PD 27-DEC-1996.

PR 07-JUN-1995; US-484722

PI Elliott KJ, Harpold MM;

DR P-PSDB; W09025.

used in screening to determine the effect of drugs on the receptor

the human neuronal nicotinic acetylcholine receptor ( $\alpha 5$ )

CC nucleic acids, opt. in combination with other alpha and/or beta

CC recombinant nACHR subunits useful for identifying cpds. that

Sequence 1876 BP; 369 A; 553 C; 531 G; 423 T;

0  
1  
2  
3  
4  
5  
6  
7  
8  
9  
A  
B  
C  
D  
E  
F  
G  
H  
I  
J  
K  
L  
M  
N  
O  
P  
Q  
R  
S  
T  
U  
V  
W  
X  
Y  
Z

Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 73; Concordance 0; Mismatches 0; 1-341-0

[illegible]

Db 1 GGC CGC AGG CGC AGG CCG GGG CGA CAC AGC CGA GAC GTG GAG CCG CGC CGC GCT C GCT GCA GCT 60

QY 381. CCGGGACTCAAC 392

Db 61 CCGGGACTCAAC 72

РЕСПУБЛИКА

standard: 1876 BP

DT 14-MAY-1998 (first entry)

KW Human; neuronal nicotinic acetylcholine receptor; alpha-7 subunit;

05 **Homo sapiens.**

	CDS	Firm
1980-1989	13.1581	144.22

```

/produced by neuronal nicotinic acetylcholine receptor
alpha-7 subunit"

```

15-SEP-1994.

PR 08-MAR-1993; US-028031

PA (SIBI-) SIBIA NEUROSCIENCES INC.









Query	Match	Similarity	Score	DB	Length
Query	Local Similarity	54.6%	14.7%	DB 1	11820
Best	Local Similarity	54.6%	Pred. No. 0.017		
Matches	183	Conservative	0	Mismatches	142
				Indels	10
				Gaps	
QY	14	GAGGTAGAGCCTTGCCCTTGGGCGACGCCCCCTGAGC--CTGGCCAGAGAGCGCGAGCCGAGAGC	71		
Db	1246	GGCGCCCGGTGGGATCATATGGCGGGCCCCCGGCTGCAGAGCGCGGGGGGGGGGGGGGGGG	1187		
QY	72	CCGCTCGGTGAGACCTGGGGGGGTGGAGGTGCCCGGAGCGTACCCAGCGCCGGGAGTACTTC	131		
Db	1186	CGGGCGCGGGGGGACGCGCGCGGGCGG-----GGATTTCACCTGGCGCGCGAGCAGCGG	1132		
QY	132	CCGCTACACACTCGGGGTGGAGTTCCTCGGTGGCGCGCCGACAGAGCGCTGGCGCGGCTGGA	191		
Db	1131	TGGCCGAGAGCGAGCGGGATCCAGACC--GGGTCCACACATGCTCATGGGGGGGGGGGGGG	1075		
QY	192	GGGATGGCGGGGCGGGAGACGGGGCGGGGCGGGGCTCTCACGTGAGAGAGCGCGCGG	251		
Db	1074	CGGGGGGGGGGGCGGGGGGGGGGGGGGGAGGGAGGGAGGGAGGGAGGGAGGGAGGGG	1015		
QY	252	GGCGGGCGGGGCGGGGCGCGCGCGGCTCTTAAAGCGCGCGAGCCGAGCGCGGAG	311		
Db	1014	CGGGCGGG	955		
QY	312	TGCTCTGTGGCCCGCAGGCCGAGGCCGGGGCGACA	346		
Db	954	AGATACGGAGCCGCTCGCGCTGATCACGGGGACA	920		
RESULT	8				
ID	X53491/c				
AC	X53491 standard; DNA; 114955 BP.				
DT	05-JUL-1999 (first entry)				
DE	Human adenosine A1 receptor antisense oligonucleotide fragment.				
KW	Antisense oligonucleotide; multiple target; antisense treatment;				
KW	impaired respiration; inflammation; lung disease;				
KW	pulmonary vasoconstriction; inflammation; allergic rhinitis;				
KW	acute asthma; allergy; asthma; impaired respiration;				
KW	respiratory distress syndrome; palm; cystic fibrosis;				
KW	pulmonary hypertension; pulmonary vasoconstriction; emphysema;				
KW	chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;				
KW	colon cancer; breast cancer; lung cancer; pancreatic cancer;				
KW	hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;				
CC	prostate cancer; ss.				
CC	Synthetic.				
CC	MO9913886-A1.				
PF	25-MAR-1999.				
PR	17-SEP-1998; U19419.				
PR	09-JUN-1998; US-093972.				
PR	17-SEP-1997; US-059160.				
PA	(UYEC.) UNIV EAST CAROLINA.				
PI	Nyce JM;				
PT	WPI; 99-229400/19.				
PT	New antisense oligonucleotides used in treatment of, e.g. pulmonary				
PT	vasoconstriction				
PS	Disclosure; Page 37; 120pp; English.				
CC	The specification describes antisense oligonucleotides (X52869-X55271)				
CC	directed against at least 2 mRNAs selected from target genes, coding and				
CC	non-coding regions of RNAs corresponding to target genes, gene				
CC	initiation codons, genomic flanking regions, intron-exon borders, the				
CC	3'-end, the 5'-end and the junction between coding and non-coding				
CC	regions and all segments of RNAs encoding proteins associated with one				
CC	or more diseases, conditions or mixtures. The antisense oligonucleotides				
CC	may be derived from sequences X55272-74. These multiple target				
CC	oligonucleotides (specifically X55180-271) can be used for the antisense				
CC	treatment of diseases and conditions. Typical diseases and conditions				
CC	are those associated with impaired respiration and inflammation,				
CC	including lung diseases, pulmonary vasoconstriction, inflammation,				
CC	allergic rhinitis, acute asthma, allergies, asthma, impaired respiration,				
CC	respiratory distress syndrome, pain, cystic fibrosis, pulmonary				
CC	hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive				
CC	pulmonary disease (COPD), and cancers such as leukemias, lymphomas,				

CC carcinoma e.g. colon cancer, breast cancer, lung cancer, pancreatic  
CC cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic  
CC metastases, as well as all types of cancers which may metastasize or have  
CC metastasized to the lungs, including breast and prostate cancer.  
SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T;

```
Query Match          14.7%; Score 57.8; DB 1; Length 114955;
Best Local Similarity 35.0%; Pred. No. 0.011;
Matches 131; Conservative 45; Mismatches 192; Indels 6; Gaps 1;

OY      22 GCGTGGCCCTGGGCGAGCCCGCTGGCCCTGCACAGAGCGCGAGAGCCGAGAGCCCGCTGGTG 81
      ||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104995 GCCCGCGCGCGCGCGCGCGCGCCCTGAGCCVGNHNNHNNNSCGGCCGCGCGCGCGCGCCVGGC 104936

OY      82 GAGACTGGGGGTGAGGTGAGTCCCGAGCAGCAGCAGCGCGCGAGTA-----CCTCCCGC 135
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104935 CVGCNNHNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGNN 104876

OY      136 TCACACTCTGGGCTGCACTCTCTGGGTGGCCCGCGAGACAGCTGAGCCCGCGGCTGAGAGGA 195
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104875 HNNNSVGGCGCGCGGNNHNNHNSCVGGCCVGGGNNHNNHNSCVGGCCVGGCGNNHNNHNSC 104816

OY      196 TGGCGGGCGGGGACGCGGGGGCGGGGGCGGGGCTGTACAGTGAAGAGCGCGCGGGGGCG 255
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104815 CCVGGCCVGGCGGNNHNNHNSGCCVGGCCVGGCGNNHNNHNSCGCCVGGCCVGGCGNNHNN 104756

OY      256 GAGCGGGCGGGGGCGCGCGCGCGCGCGGCTCTTAAAGCGCGCGAGCCGAGCGGAGGTGCC 315
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104755 NSGCGCCCGVGGCGVGGCGGNNHNNHNSCGCGCGCGCGCGCGCGCGGNNHNNHNSCGCGCCVGG 104696

OY      316 TCTGTGGCCGAGCGAGCGCGCGCGCGCGAGACAGCAGTGTGAAGCGCGCGCGCTGCTG 375
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104695 CCVGGCGGNNHNNHNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104636

OY      376 CAGCTCCGGGACTC 389
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104635 NNNHNNHNSGCGCGCGC 104622

RESULT 9
ID      9 076213 standard; DNA; 12001 BP.
AC      076213;
DR      04-AUG-1995 (first entry)
DE      HSV L/ST region.
KW      HSV; junction-spanning transcript; L/ST; therapy; virucide; ds.
OS      Herpes simplex virus type 1.
FT      Key Location/Qualifiers
FH      tata-signal 199..204
FT      /tag= a
FT      misc_binding 220..225
FT      /tag= b
FT      /note= "ICP4 binding site"
FT      misc_rna 229
FT      /tag= c
FT      /note= "5' end of the L/STs"
FT      misc_rna 370..372
FT      /tag= d
FT      /note= "first codon of ORF-1"

PN      WO9428156-A.
PD      08-DEC-1994.
PE      20-MAY-1994; U05770.
PR      20-MAY-1993; US-065146.
PA      (DAND ) DANA FARBER CANCER INST INC.
PI      Schaffer PA, Yeh L;
PT      Herpes Simplex Virus (HSV) specific junction spanning transcript
PT      - for inhibiting HSV L/ST synthesis, in the treatment of HSV
PT      infection.
PS      Disclosure; Page 38-44; 64pp; English.
CC      An HSV-specific junction-spanning transcript (L/ST) maps at the 5'
CC      end to the b repeat sequences of HSV DNA at approx. 3 and 125 kb,
```

CC and extends into the c repeat sequences of HSV DNA. Compounds  
 CC that inhibit HSV L/ST synthesis may be used as anti-HSV  
 CC virucides. The HSV-1 DNA sequence in the region of the L/STs  
 CC is given in 076213  
 SQ Sequence 12001 BP; 1568 A; 4557 C; 4315 G; 1561 T;

Query Match 14.4%; Score 56.4; DB 1; Length 12001;  
 Best Local Similarity 50.3%; Pred. No. 0.028;  
 Matches 164; Conservative 0; Mismatches 101; Indels 1; Gaps 1;

QY 43 GGCCTGCGCAGAGCGCGAGCGCGAGCGCGCTCGTGAGACTGGGGGTGAGGTGCC 102  
 DB 2319 GGGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGG 2260  
 QY 103 CGGAGCGTACCGAGCGCGCGAGTACCTCCGCTCACACCTTCGGCTCACTTCCCTGGG 162  
 DB 2259 CGAGGGGGGGGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGG 2200  
 DB 163 TGGCCGCGGAGACGCTGGCCCGGCTGGAGGAGATGGCGGGGCGAGGGGGGGGCG 222  
 QY 2139 GAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGG 2140  
 QY 223 GGGGCTGCTACGCTGAGAGAGCGCGCGGGGGCGCGGGGGCGCGCGCGCGCTC 282  
 DB 2139 GAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGG 2080  
 QY 283 CTTAAGCGCGCGAGCGCGAGCGCGAGTCTCTGTGCGCGCAGCGCGAGCGCGCGG 342  
 DB 2079 GAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGG 2021  
 QY 343 GACAGCGGAGACGTGAGCGCGCGCGG 368  
 DB 2020 GAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGG 1995

RESULT 10  
 V60339  
 ID V60339 standard; cDNA; 1120 BP.  
 AC V60339;  
 DT 04-DEC-1998 (first entry)  
 KW cDNA sequence of fibroblast growth factor-2 (FGF-2).  
 KW Fibroblast growth factor-2; FGF-2; leaderless protein; inhibition;  
 KW export; angiogenesis; restenosis; treatment; tumour; inflammation;  
 KW cell proliferation; diabetes; retinopathy; infection;  
 KW polycystic kidney disease; atherosclerosis; ss.  
 OS Homo sapiens.  
 WO9837880-A1.  
 03-SEP-1998.  
 25-FEB-1998; U03689.  
 26-FEB-1997; US-807014.  
 PA (CIBL) CIBLEX CORP.  
 PI Baird A, Florkiewicz R;  
 WT: 98-495377/42.  
 DR Inhibiting export of leaderless protein with agent that inhibits  
 PT binding to transporter protein - especially for treating  
 PT angiogenesis and restenosis by preventing export of fibroblast  
 PT growth factor, also methods for identifying leaderless proteins and  
 PT their transporters  
 PS Claim 2; Pages 53-54; 116pp; English.  
 CC The present sequence encodes fibroblast growth factor-2 (FGF-2), a  
 CC leaderless protein. A leaderless protein refers to a protein that is  
 CC found in an extracellular environment, but lacks a canonical leader  
 CC sequence. The specification describes a method for inhibiting export of  
 CC a leaderless protein from a cell. The method comprises treating  
 CC the cell with an agent that inhibits binding between the leaderless  
 CC protein and a transport molecule. Treatment with the inhibiting agent  
 CC is specifically used to treat angiogenesis and restenosis, i.e. where  
 CC expression of FGF-2 is inhibited, and the agent is applied to endothelial  
 CC or smooth muscle cells. Other applications are treatment of tumours  
 CC (melanoma, teratocarcinoma, ovarian carcinoma, bladder cancer and  
 CC neuroblastoma), inflammation, cell proliferation, complications of  
 CC diabetes (e.g. retinopathy), viral, bacterial or fungal infections,

CC polycystic kidney disease and atherosclerosis.  
 SQ Sequence 1120 BP; 260 A; 293 C; 342 G; 225 T;  
 Query Match 14.3%; Score 56; DB 1; Length 1120;  
 Best Local Similarity 47.7%; Pred. No. 0.052;  
 Matches 164; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 41 CTGGCCCTGCGCAGAGCGCGAGCGCGAGCGCGCTCGTGAGACTGGGGGTGAGAGTG 100  
 DB 83 CGGCTCTCCCAAGCGCGCGCTCCCGAGACACCATCTGTAACCCCAAGTCCCGGGCG 142  
 QY 101 CCGGAGCGTACCGAGCGCGCGAGTACCTCCGCTCACACCTTCGGGCTCACTTCCCTG 160  
 DB 143 CCGGCTGCGCGCGCGAGCGCGCGAGGGGGCGAGAGAGAGGGGGCGAGCGCGCTGAGGCTG 202  
 QY 161 GGTGGCCCGCGAGACGCTGGCCCGGCTGGAGGAGTGGCGGGGGCGAGCGGGGGG 220  
 DB 203 GGGGACCGCGCGCGCGCGCGCTCCCGGCGGAGGCGCTGGGGGGCGCGGGGGG 262  
 QY 221 GCGGGGCTGCTACGCTGAGAGAGCGCGCGGGGGCGGGGGGGCGGGGGCGCGCGCGG 280  
 DB 263 CCGGCGCGCGAGCGCGCTCGAGAGCGCGGGCGCGGGGGCGAGCGCGCTCCCGCGG 322  
 QY 281 TCGTTAAGCGCGCGAGCGCGAGCGCGAGTCTCTGTGCGCGCGAGCGCGAGCGCGG 340  
 DB 323 CGGCTCCAGCGCGCTCGGGGATCCCGCGGGGGCGCGCGAGGACCATGCGAGCGGGAGCA 382  
 QY 341 GCGACAGCGGAGACGTGAGCGCGCGCGCTCGCTGCACTCCG 384  
 DB 383 TCACACGCTGCTCCGCTTGCGCGAGATGGGGCGAGCGCGG 426

RESULT 11  
 Q76213  
 ID Q76213 standard; DNA; 12001 BP.  
 AC Q76213;  
 DT 04-AUG-1995 (first entry)  
 DE HSV L/ST region.  
 KW HSV; junction-spanning transcript; L/ST; therapy; virucide; ds.  
 OS Herpes simplex virus type 1.  
 FH Key location/Qualifiers  
 FT 199..204  
 FT /\*tag- a  
 FT misc\_binding 220..225  
 FT /\*tag- b  
 FT /\*note- "ICP4 binding site"  
 FT misc\_rna 229  
 FT /\*tag- c  
 FT /\*note- "5' end of the L/STs"  
 FT misc\_rna 370..372  
 FT /\*tag- d  
 FT /\*note- "first codon of ORF-1"

PN WO9428156-A;  
 PD 08-DEC-1994.  
 PE 20-MAY-1994; U05770.  
 PR 20-MAY-1993; US-065146.  
 PA (DAND) DANA FARBER CANCER INST INC.  
 PI Schaffer PA, Yen L;  
 DR WI; 95-022825/03.  
 PT Herpes Simplex Virus (HSV) specific junction spanning transcript  
 PT - for inhibiting HSV L/ST synthesis, in the treatment of HSV  
 PT infection.  
 PS Disclosure, Page 38-44; 64pp; English.  
 CC An HSV-specific junction-spanning transcript (L/ST) maps at the 5'  
 CC end to the b repeat sequences of HSV DNA at approx. 3 and 125 kb,  
 CC and extends into the c repeat sequences of HSV DNA. Compounds  
 CC that inhibit HSV L/ST synthesis may be used as anti-HSV  
 CC virucides. The HSV-1 DNA sequence in the region of the L/STs  
 CC is given in 076213.  
 SQ Sequence 12001 BP; 1568 A; 4557 C; 4315 G; 1561 T;

OY	173	GACGCTGGCCCCCGGTGTGAAGGATATGCAGCGGCGGAGACGGGGCGTGGGGCCGGGCTGTC	232
Db	253	CCCCCCTCCCCTCCCGGTGTGTGTGGTGCGGCGGCCCGCCCGGCGCGGCGCCGGCCGG	194
OY	223	ACGTGAGAAGCGCGCGGGGGCGGGGGCGGGGGCGCGCGCCCGGCTCTTAAGCGC	292
Db	193	CGCAGCGCGCGCGCGCGGAGGCAATATGCGGCACACAGGCTCGCG--CGGGCCCGAAGCTCG	136
OY	293	CGCAGCGCGAGCGCGGAGAGTGCCTCTGTGCGCGCACGCGCGACAGGCCCGGCGACAGCCGAG	352
Db	135	AGTGCGCGGGGGGGGGAGTAGTGCGCGCTTTAAAGCCGCTCGCGCGCGGGAAGCGCGCC	76
OY	353	ACGTGAGACCGCGCGCGCTGCTGCAGCTCCGG	385
Db	75	GCCGCGCCACC CGCGCGCCGACACCCCGCG	43
<b>RESULT_13</b>			
ID	Q47929/c	standard; cDNA; 2218 BP.	
AC	Q47929;		
DT	24-MAR-1994	(first entry)	
KW	Paired basic amino acid converting enzyme (PACE) 4.1 gene.		
KV	Paired basic amino acid converting enzyme; PACE; prorenin; renlin; hypertension; blood; coagulation; ss.		
OS	Homo sapiens.		
FH	Key	Location/Qualifiers	
FT	cds	/tag= a	
FT		/product= PACE 4.1.	
PN	WO9318159-A.		
PD	16-SEP-1993.		
PF	09-MAR-1993; U02147.		
PR	09-MAR-1992; US-848629.		
PA	(CHIR ) CHIRON CORP.		
DR	Barr PJ, Kiefer MC; WPI; 93-303473/38.		
P	P-PSDB; R41663.		
PT	New polynucleotide encoding PACE endo-peptidase - used for e.g. reducing blood pressure		
PS	Example 2; Figures 1 and 3; 81pp; English.		
CC	The mammalian endopeptidases PACE 4 and PACE 4.1 (an alternative form of PACE 4) are involved in the production of mature poly peptides from precursor polypeptides by cleavage at pairs of basic amino acids e.g Lys-Arg, Lys-Lys and Arg-Arg. The enzymes are stimulated by the presence of calcium ions and inhibited by phenylmethyl sulphonyl fluoride. PACE 4 and 4.1 convert prorenin to renin in the human kidney, causing elevation of blood pressure. They can therefore be used to identify other inhibitors of their action and may also be used to inhibit blood coagulation.		
SC	Sequence 2218 BP; 534 A; 623 C; 623 G; 438 T;		
<b>Query Match</b> 13.2%; Score 51.8; DB 1; Length 2218; Best Local Similarity 49.2%; Pred. No. 0.21; Mismatches 167; Indels 2; Gaps 1;			
Matches	164; Conservative	0; Mismatches	167; Indels 2; Gaps 1;
OY	53	GAGCGCGAGGCGCGAGACCCGCTCGGTGAGACTGGGGGTGAGAGTGC CGGAGCGTAC	112
Db	373	GAGGCGCGCGCGGGGCGCGGAGCAGGCGGCGAGCGCCAGCGACACAGCGCGCA	314
OY	113	CCAGGCGCGGAGTAGTACCTCCCGCTCACACCTCGGGGCTCAATTCCCTGAGTGGCGCGGA	172
Db	313	GGGAGCGCGCGCGAGCGCGCGGACCCCGGCGCGCGCGCGCGCGCGCGCGCGCC	254
OY	173	GACGCTGGCGCGCGGTGAGGAGATGGACGCGGCGGAGAGGGGCGGGGCGGCTGTC	232
Db	253	CCCCCGCGCGCGCGGTGTGTGTGGTGCGGCGGCCCGCCCGGCGCGGCGCCGG	194
OY	223	ACGTGAGAAGCGCGCGGGGGCGGGGGCGGGGGCGCGCGCCCGGCTCTTAAGCGC	292
Db	193	CGCAGCGCGCGCGCGCGGAGGCAATATGCGGCACACAGGCTCGCG--CGGGCCCGAAGCTCG	136
OY	293	CGCAGCGCGAGCGCGGAGAGTGCCTCTGTGCGCGCACGCGCGACAGGCCCGGCGACAGCCGAG	352
Db	135	AGTGCGCGGGGGGGGGAGTAGTGCGCGCTTTAAAGCCGCTCGCGCGCGGGAAGCGCGCC	76
OY	353	ACGTGAGACCGCGCGCGCTGCTGCAGCTCCGG	385
Db	75	GCCGCGCCACC CGCGCGCCGACACCCCGCG	43

QY 293 CCGGAGCCGAGGCGAGGTGCTGTGCGCCGAGCGGAGCCCGGCGACACCCAG 352  
 DB 135 ACTGGCCCGGGGGTGGATGCGCTTTAAAGCCGCTCCGCGGGGAGCGCCGCC 76  
 QY 353 ACGTAGAGCGCGCGCTGCTGACAGTCCGGG 385  
 DB 75 GCGCGCGCCACCGCGCGCCGACACACCGCCGG 43  
 RESULT 14  
 ID V10362/c  
 AC V10362: standard; cDNA; 4257 BP.  
 DE 30-JUN-1998 (first entry)  
 DT Infected cell protein number 4 alpha-4 gene.  
 KW Infected cell protein number 4; ICP4; alpha-4; cell apoptosis;  
 OS Herpes simplex virus.  
 FH Key Location/Qualifiers  
 CDS 361..4257  
 /tag= a  
 /product= "Infected cell protein"

PN M09804709-A2.  
 PD 05-FEB-1998.  
 PF 23-JUL-1997; U12904.  
 PR 26-JUL-1996; US-690473.  
 PA (ARCH-) ARCH DEV CORP.  
 PI Leopardi R, Roizman B;  
 DR WPI: 98-130697/12.  
 P-PSDB: M40200.  
 PT Use of herpes simplex virus ICP4 polypeptide - useful for, e.g.  
 PR blocking apoptosis in cells, production of proteins and gene therapy  
 CC The herpes simplex virus (HSV) alpha-4 gene encodes infected cell protein  
 CS number 4 (ICP4). Both the alpha-4 gene and ICP4 can be used to block  
 CC cell apoptosis. Similarly the administration of an agent that inhibits  
 CC ICP4 or the alpha-4 gene can induce apoptosis in HSV infected cells.  
 CC This can be used for the immortalisation of cells, production of  
 CC proteins, gene therapy, or inhibition of cell death induced in vivo.  
 CC They can also be used for production of therapeutics comprising  
 CC inhibitors of HSV ICP4 function, useful for treating HSV function.  
 SQ Sequence 4257 BP; 412 A; 1768 C; 1663 G; 414 T;

Query Match 13.2%; Score 51.8; DB 1; Length 4257;  
 Best Local Similarity 46.5%; Pred. No. 0.19;  
 Matches 167; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

DB 27 GCCTTGGGAGCCCTGCGCTGCGCCAGAGCGCGGAGAGCCCGCTCGGTGAGAC 86  
 DB 2806 GCCGCGACCGCCGAGGGGGTGGGGCCCTCGCGGGCGCGGACACAGCGCCAGGGGC 2747  
 QY 87 TGGGGGTGAGGTGCGCCGAGCGTACCCAGCCGCGGAGTACCTCCGCTACACCTCGG 146  
 DB 2746 GGGGGGGGCGCTGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 2687  
 QY 147 GCTGAGTTCCTTGGGTGCGCCGAGAGCGTGGCCCGGGCTGAGGAGTGGCGGGGGCG 206  
 DB 2686 GCGCCGCGGGGCTGAGAGGGGGGCGGGGAGGGGGGCGGGGCGCTCGAGAGCGG 2627  
 QY 207 GGACGG 266  
 DB 2626 GGGCGTCCGCGCGCTCTTCTTCTTCTTGGGGGTGCGGGGCGCGCGCTCGGGCGGGC 2567  
 QY 267 GGGCGCGCGCGCGCTCTTAAAGCGCGAGCGAGCGCGAGGTGCGCTCTGTGGCGCG 326  
 DB 2566 GGGCGCGCGCGGAGCTTCTGCGCTTGGCGCCCTCCGCGGGGGGGGGGGGGGGGG 2507  
 QY 327 AGGCGAGCGCGCGGAGAGCGTGAAGCGCGCGCGCTGCTGCACTCCGGG 385  
 DB 2506 CCGCAGCGCGTGGGGGGGGTGGCGTGGCGCGCGCGCGCAGAGGGGGGGCGCAGG 2448

RESULT 15  
 ID V68520/c  
 AC V68520:  
 DT 29-JAN-1999 (first entry)  
 DE The nucleotide sequence of the Herpes simplex virus ICP4.  
 KW Herpes simplex virus ICP4; HSV ICP4; modulator; apoptosis;  
 OS stimulation; inhibition; HSV infection; ss.  
 FH Key Location/Qualifiers  
 CDS 361..4257  
 /tag= a  
 /product= "HSV ICP4 protein"

PN M09846637-A2.  
 PD 22-OCT-1998.  
 PF 16-APR-1998; U07573.  
 PR 16-APR-1997; US-843659.  
 PA (ARCH-) ARCH DEV CORP.  
 PI Leopardi R, Roizman B;  
 DR WPI: 98-594559/50.  
 P-PSDB: W80810.  
 PT Use of herpes simplex virus U(s)3 polypeptide - for developing  
 PR products for modulating apoptosis in cells and for identifying  
 PT compounds which act as stimulators or inhibitors of apoptosis  
 PS Example 2, Pages 60-63; 85pp; English.  
 CC This is the nucleotide sequence of Herpes simplex virus ICP4 used in  
 CC the method of the invention as modulators of apoptosis. The methods  
 CC and products can be used to identify compounds which modulate  
 CC (stimulate or inhibit) apoptosis in cells. They can be used to  
 CC immortalise cells for the study of these cells or for growing cells  
 CC in large numbers for the productions of proteins. They can also be  
 CC used for stimulating apoptosis in cells, e.g. for treating a subject  
 CC with a HSV infection.  
 SQ Sequence 4257 BP; 412 A; 1768 C; 1663 G; 414 T;

Query Match 13.2%; Score 51.8; DB 1; Length 4257;  
 Best Local Similarity 46.5%; Pred. No. 0.19;  
 Matches 167; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 27 GCCTTGGGAGCCCTGCGCTGCGCCAGAGCGCGGAGAGCCCGCTCGGTGAGAC 86  
 DB 2806 GCCGCGACCGCCGAGGGGGTGGGGCCCTCGCGGGCGCGGACACAGCGCCAGGGGC 2747  
 QY 87 TGGGGGTGAGGTGCGCCGAGCGTACCCAGCCGCGGAGTACCTCCGCTACACCTCGG 146  
 DB 2746 GGGGGGGGCGCTGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 2687  
 QY 147 GCTGAGTTCCTTGGGTGCGCCGAGAGCGTGGCCCGGGCTGAGGAGTGGCGGGGGCG 206  
 DB 2686 GCGCCGCGGGGCTGAGAGGGGGGCGGGGAGGGGGGCGGGGCGCTCGAGAGCGG 2627  
 QY 207 GGACGG 266  
 DB 2626 GGGCGTCCGCGCGCTCTTCTTCTTCTTGGGGGTGCGGGGCGCGCGCTCGGGCGGGC 2567  
 QY 267 GGGCGCGCGCGCGCTCTTAAAGCGCGAGCGAGCGAGGTGCGCTCTGTGGCGCG 326  
 DB 2566 GGGCGCGCGCGGAGCTTCTGCGCTTGGCGCCCTCCGCGGGGGGGGGGGGGGGGG 2507  
 QY 327 AGGCGAGCGCGCGGAGAGCGTGAAGCGCGCGCGCTGCTGCACTCCGGG 385  
 DB 2506 CCGCAGCGCGTGGGGGGGGTGGCGTGGCGCGCGCGCGCAGAGGGGGGGCGCAGG 2448

Search completed: September 18, 1999, 00:34:58  
 Job time: 19058 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:05:04; Search time 1962.17 Seconds

(without alignments)  
1116.740 Million cell updates/sec

Title: US-08-956-518a-102

Sequence: 1 AGCCTTCCAGGCGGTAG.....CAAGATTCTGCTACATCG 689

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

abase :

GenEmb1:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_om:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_ph:\*  
8: gb\_ph:\*  
9: gb\_ph:\*  
10: gb\_ph:\*  
11: gb\_ph:\*  
12: gb\_ph:\*  
13: gb\_ph:\*  
14: gb\_ph:\*  
15: gb\_ph:\*  
16: gb\_ph:\*  
17: gb\_ph:\*  
18: gb\_ph:\*  
19: gb\_ph:\*  
20: gb\_ph:\*  
21: gb\_ph:\*  
22: gb\_ph:\*  
23: gb\_ph:\*  
24: gb\_ph:\*  
25: gb\_ph:\*  
26: gb\_ph:\*  
27: gb\_ph:\*  
28: gb\_ph:\*  
29: gb\_ph:\*  
30: gb\_ph:\*  
31: gb\_ph:\*  
32: gb\_ph:\*  
33: gb\_ph:\*  
34: gb\_ph:\*  
35: gb\_ph:\*  
36: gb\_ph:\*  
37: gb\_ph:\*  
38: gb\_ph:\*  
39: gb\_ph:\*  
40: gb\_ph:\*  
41: gb\_ph:\*  
42: gb\_ph:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	689	100.0	689	11	AF029839	Homo sapi

2	503	73.0	641	11	AF029838	AF029838 Homo sapi
3	292.6	42.5	137699	11	AC005630	AC005630 Homo sapi
4	282.2	41.0	291	9	HS2E10F	Z56126 H.sapiens C
5	269.4	39.1	192439	35	AC007562	AC007562 Homo sapi
6	261	37.9	178307	35	AC006328	AC006328 Homo sapi
7	237.2	34.4	298	9	HS2G7F	Z60646 H.sapiens C
8	155.4	22.6	1712	11	AF036903	AF036903 Homo sapi
9	154.4	22.4	1836	11	AF037646	AF037646 Homo sapi
10	118.8	17.2	180	9	HS1562F	Z54415 H.sapiens C
11	116.2	16.9	171	9	HS1651F	Z57156 H.sapiens C
12	108	15.7	1559	9	HUM47NAR	L25827 Human a7 n1
13	108	15.7	2087	10	HSAR47A	X70297 H.sapiens m
14	108	15.7	1509	10	HSNACHA7	Y08420 H.sapiens m
15	108	15.7	1876	10	HSU62436	U62436 Human n1cot
16	106.4	15.4	1555	10	HSCHRNA7A	Z23141 H.sapiens C
17	106.4	15.4	1977	10	HSU40583	U40583 Human alpha
18	93	13.5	5281	9	HSU08191	U08191 Human R kap
19	92	13.4	1848	12	MDSNARS	L37663 Mus musculu
20	90.4	13.1	2106	12	RATNARAD	I31619 Rattus rat
21	90.4	13.1	3030	12	S53987	S53987 nicotinic x
22	89.4	13.0	139	9	HS156G10R	Z64674 H.sapiens C
23	88.8	12.9	1551	3	BN7A7	X93604 B.taurus mr
24	74.4	10.8	2696	4	GGA7NAREC	X52295 Chicken alp
25	74.4	10.8	2037	4	GGNNA7A7A	X65586 G.gallus mr
26	61.6	8.9	2090	4	GGABNAREC	X52296 Chicken alp
27	60.4	8.8	189	9	HS15D8R	Z61392 H.sapiens C
28	58.2	8.4	3029	37	AF143847	AF143847 Helicobis
29	50	7.3	89871	37	AC005890	AC005890 Drosophill
30	48.2	7.0	32039	1	SC7A1	AL034447 Streptomy
31	46.4	6.7	336	14	G47584	G47584 Z2357_1 Ze
32	46.2	6.7	76947	36	AC004326	AC004326 Drosophill
33	46.2	6.7	3629	37	AF143846	AF143846 Helicobis
34	44.6	6.5	2493	12	RNNIACERE	X15834 R.norvegicu
35	44.6	6.5	2461	12	RNNI42976	U42976 Rattus norv
36	44	6.4	1008	12	RATNACHRRS	M33952 Rat neutroa
37	42.8	6.2	143664	35	AC007291	AC007291 Drosophill
38	42.6	6.2	1574	11	AF069682	AF069682 Homo sapi
39	42.4	6.2	3459	10	HS012582	AF012582 Homo sapi
40	42.4	6.2	785	10	HS0133727	AJ133727 Homo sapi
41	42.2	6.1	107889	11	AC004519	AC004519 Homo sapi
42	42.2	6.1	42301	11	AC005943	AC005943 Homo sapi
43	41.6	6.0	120766	42	AC004150	AC004150 Homo sapi
44	41.2	6.0	2672	3	RAT7HBP	J05602 Rabbit mult
45	41.2	6.0	4085	4	CHKACHR02	K02904 Chicken nic

## ALIGNMENTS

RESULT	1	AF029839	689 bp	MRNA	PRI	16-DEC-1998
LOCUS	AF029839					
DEFINITION	Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.					
ACCESSION	AF029839					
NID	93757794					
VERSION	AF029839.1	GI:3757794				
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.					
AUTHORS	Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J., Moore, T., Jacobs, S., Meriwether, J., Choi, M., Kim, E., J., Walton, K., Butting, K., Davis, A., Breese, C., Freedman, R., and Leonard, S.					
TITLE	Genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene					
JOURNAL	Genomics 52 (2), 173-185 (1998)					
MEDLINE	99000837					
REFERENCE	2 (bases 1 to 689)					
AUTHORS	Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R., Breese, C., Davis, A., Hopkins, J., and Freedman, R.					
TITLE	Direct Submission					

JOURNAL Submitted (15-OCT-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA

# FEATURES

source  
1. 689  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15q14"  
/clone="CHRNA7-DR2"  
/note="RACE product A/C/D"  
/note="alpha 7 neuronal nicotinic receptor mRNA; used in alternative splicing"  
misc\_feature  
BASE COUNT 142 a 210 c 196 g 141 t  
ORIGIN

Query Match 100.0%; Score 689; DB 11; Length 689;  
Best Local Similarity 100.0%; Pred. No. 2e-156;  
Matches 689; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AGCCCTTCCAGCGGTAAGCGGGGAGTGTGTCTGTTGCCCTTTAACTGCGGCTTG 60  
1 AGCCCTTCCAGCGGTAAGCGGGGAGTGTGTCTGTTGCCCTTTAACTGCGGCTTG 60  
61 ACGGAGACCGCGCTCTCTGCTGCTGAGTGGTTATTAAGAGACGCCCGCAGCGCCG 120  
121 CACATAGCTCCCGCAAGTCTCGGTGCGCTTCCATTTCCAGCGCGGCTCCCAAGAG 180  
121 CACATAGCTCCCGCAAGTCTCGGTGCGCTTCCATTTCCAGCGCGGCTCCCAAGAG 180  
181 GGTACAGCGCGCGGAGAGAGTGAAGCGCGAGAGCTCGCGCGGCGCGCGCTGCTG 240  
181 GGTACAGCGCGCGGAGAGTGAAGCGCGAGAGCTCGCGCGGCGCGCGCTGCTG 240  
241 CCGCGGCGATGACAGAGGCGCTCGGAGTGGCTTTTCCGCGCGCGCGGAGGTG 300  
241 CCGCGGCGATGACAGAGGCGCTCGGAGTGGCTTTTCCGCGCGCGCGGAGGTG 300  
301 AGGGAAGATGTCATGTGAGGTTCAAGGCAACGGAAGTACTGGCTTATCTTCC 360  
301 AGGGAAGATGTCATGTGAGGTTCAAGGTTCAAGGCAACGGAAGTACTGGCTTATCTTCC 360  
361 AGGGAAGATGTCATGTGAGGTTCAAGGTTCAAGGCAACGGAAGTACTGGCTTATCTTCC 360  
361 AGGGAAGATGTCATGTGAGGTTCAAGGTTCAAGGCAACGGAAGTACTGGCTTATCTTCC 360  
421 GACACCTGAGTCAAGAGAGCTTGAATCCAGATGAGAGGCTTATCTTACAGACTCA 480  
421 GACACCTGAGTCAAGAGAGCTTGAATCCAGATGAGAGGCTTATCTTACAGACTCA 480  
421 GACACCTGAGTCAAGAGAGCTTGAATCCAGATGAGAGGCTTATCTTACAGACTCA 480  
481 TCTTGTGTACACCCCATTTATGACAATCCAAAGGTGAGAAAGCACTGGAATTTCA 540  
481 TCTTGTGTGTACACCCCATTTATGACAATCCAAAGGTGAGAAAGCACTGGAATTTCA 540  
541 ATTGCTAATCCAGCATTTGTGATAGTGAAGTGAATCTGATGATGAGAGGCTTTGA 600  
541 ATTGCTAATCCAGCATTTGTGATAGTGAAGTGAATCTGATGATGAGAGGCTTTGA 600  
601 CCGCACAATTCACACTAAGCTGTGTGTAATTTCTTGGGCAATGACAGTACTGCTCC 660  
601 CCGCACAATTCACACTAAGCTGTGTGTAATTTCTTGGGCAATGACAGTACTGCTCC 660  
661 AGGCAATTTCAAGAGTTCCTGTAACATCG 689  
661 AGGCAATTTCAAGAGTTCCTGTAACATCG 689

RESULT 2  
AF029838 AF029838 641 bp MRNA PRI 16-DEC-1998  
LOCUS

DEFINITION Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.  
ACCESSION AF029838  
NID 93757793  
VERSION AF029838.1 GI:3757793  
KEYWORDS  
SOURCE  
ORGANISM human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.

# REFERENCE

1 (bases 1 to 641)  
Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J.,  
Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K.,  
Bulting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.  
Genomic organization and partial duplication of the human alpha7  
neuronal nicotinic acetylcholine receptor gene  
Genomics 52 (2), 173-185 (1998)

# TITLE

2 (bases 1 to 641)  
Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,  
Breese, C., Davis, A., Hopkins, J. and Freedman, R.  
Direct Submission

# JOURNAL

Submitted (15-OCT-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA

# FEATURES

source  
1. 641  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15q14"  
/clone="CHRNA7-DR1"  
/note="RACE product A/B/C/D"  
/note="alpha 7 neuronal nicotinic receptor mRNA; used in alternative splicing"  
misc\_feature  
BASE COUNT 151 a 192 c 163 g 135 t  
ORIGIN

Query Match 73.0%; Score 503; DB 11; Length 641;  
Best Local Similarity 90.0%; Pred. No. 1.2e-111;  
Matches 577; Conservative 0; Mismatches 0; Indels 64; Gaps 1;

113 CAGGCGGCGCATCTCTCCGCAAGTCTGCTGCTTCCATTTTCCAGCGCGCT 172  
1 CAGGCGGCGCATCTCTCCGCAAGTCTGCTGCTTCCATTTTCCAGCGCGCT 172  
173 CCCAGAGAGGTACGCGCGGAGAGTGAAGCGCGAGAGCTCGCGCGGCGCGCG 232  
61 CCCAGAGAGGTACGCGCGGAGAGTGAAGCGCGAGAGCTCGCGCGGCGCGCG 120  
233 CCGTGTGCGCGGCGCATCTCTCCGCAAGTCTGCTGCTTCCGCGCGCTCCCGC 292  
121 CCGTGTGCGCGGCGCATCTCTCCGCAAGTCTGCTGCTTCCGCGCGCTCCCGC 180  
233 CCGAGTGAAGGGAAGATGTCATGTCAAGGTTCAAGGCAACGGAATTTACTGGCTC 352  
181 CCGAGTGAAGGGAAGATGTCATGTCAAGGTTCAAGGCAACGGAATTTACTGGCTC 240  
353 TATCTTCCAGAGGAACGAGGAGCCAGCGCGGCTCAAGCGCCCAACGCAATTAAGT 412  
241 TATCTTCCAGAGGAACGAGGAGCCAGCGCGGCTCAAGCGCCCAACGCAATTAAGT 300  
413 TACAAGTGAACCTGAGTCAAGAGTGAATCCAGATGAGAGAGCTTATCTACAC 472  
301 TACAAGTGAACCTGAGTCAAGAGTGAATCCAGATGAGAGAGCTTATCTACAC 360  
473 GACTAGATCTTGTGTCAACCCCAATTTATGACAATCCAAAGGTCCAGAAAGCTCTGA 532  
361 GACTAGATCTTGTGTCAACCCCAATTTATGACAATCCAAAGGTCCAGAAAGCTCTGA 420  
533 CAA-----  
421 CAAATAATGAACAACCAACCATCGGTTAAATTTGATGCAAAAATTTGCAATCTACACGA 480



```

OY 536 ----TTCACATGCAATATCCACCATTTGGGAGTAGCTGCAACTGGCAATTTGCTTA 588
Db 481 TTTTCAGTTCATTCATTCGTAATCCACCATTTTGCGATTAGCTGCAGAACTGCGAATTTGCTTA 540
OY 589 TGAGGCGCTTTGACGCGACATTCACACATCAAGCTGTGTGGTAATTCCTTCGGGCATTTGCCA 648
Db 541 TGAGGCGCTTTGACGCGACATTCACACATCAAGCTGTGTGGTAATTCCTTCGGGCATTTGCCA 600
OY 649 GTACCTGCCTCCAGGCATATTCAGAGTTCTCTGCTACATCG 689
Db 601 GTACCTGCCTCCAGGCATATTCAGAGTTCTCTGCTACATCG 641

RESULT 3
AC005630
LOCUS AC005630 137699 bp DNA PRI 15-JAN-1999
DEFINITION Homo sapiens clone Dll129D05, complete sequence.
VERSION AC005630
KEYWORDS AC005630.1 GI:4159882
SOURCE HTG.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Euthelia; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 137699)
TITLE Waterston,R.H.
JOURNAL The sequence of Homo sapiens clone
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 137699)
TITLE Waterston,R.H.
JOURNAL Direct Submission
AUTHORS Submitted (08-SEP-1998) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 137699)
AUTHORS Waterston,R.H.
JOURNAL Direct Submission
AUTHORS Submitted (15-JAN-1999) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jan 15, 1999 this sequence version replaced g1:3550965.
FEATURES
source
Location/Qualifiers
1..137699
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Dll129D05"
BASE COUNT 36908 a 30670 c 32412 g 37709 t
ORIGIN

```

	Query Match	42.5%	Score 292.6	DB 11	Length 137699
	Best Local Similarity	83.1%	Pred. No. 6.5e-61		
	Matches 345	Conservative	0	Mismatches 69	Indels 1
				Gaps	1
OY	1	AGCCCTTCCAGCGCGTAGCGGGGCGAAGTGCTGTGCCCTTTAAACTGCGGCTTG	60		
Db	47346	AGCCCTTCCAGCGCGTAGCGGGGCGTGTGCTGTGCCCTTTAAAGTGGCGCTTG	47405		
OY	61	ACGGAGCGCGGCCCTCGTGGTGGAGTGGGTTTAAAGGAGCGCCCGCAGGCGCG	120		
Db	47406	ACAGGAGCGCGGCCCTCGTGGTGGAGTCTTTACAAAGGAGCGCGCCCGCAGGCGCG	47465		
OY	121	CACATAGCT-CCGCGCAAGTCTCGGTGGGCCCTTGCCATTTTCCAGCGCGCTCCACAGA	179		
Db	47466	CACACAGCTCCCGCGAGAGGCTCGGTGGGCCCTTGCCATTTTCCAGCGCTCCACAGTA	47525		
OY	180	GGGTACAGCGCGCGGGGGAGAGGTGGAGCGCGAGAGCTGGCGCGGGGGCGCCGCGCTGGG	239		
Db	47526	GAGTGGAGGCACTCAGGAGAGAGCGCGGAGCTGGGAGAGAGCGCCCGCGAGAGTCCCGCGGGTG	47585		
OY	240	GCCGCGGCCATGACAGCGGCTCGGGAGCTGGCTCTTTTCCGCGCCCTTCGCGCCGAGGT	299		

Db	47586	GTGGCGGCGGTGACAGCGGCTCCGAGCGGCTCAACTTCGGCGCCCTCCCGCCAGAGGT	47645
Oy	300	GAGGGGAGATGTCCATGTGCAAGGTTCAAGGCCAAACGAGTTACTGGCCTTATCTTC	359
Db	47646	GAGGTAAATGTCCCTGTGAGGGTTCAAGGCCAAGGTGAGTGTGGCCTTATCTTC	47705
Oy	360	CAGAGACACGAGAGCCACAGCGCGGCTACGCCGCCACCGCAATTAGATT	414
Db	47706	CACAGAACCGAGAGCGCGCGCGCCGACGTCTCCACTGCACATCAGGTTGA	47760

RESULT	4
HS2E10F	
LOCUS	HS2E10F 291 bp DNA PRI 18-Oct-1995
DEFINITION	H.sapiens Cpg Island DNA genomic MseI fragment, clone 2e10, forward read cp92e10.f.tild.
ACCESSION	Z58126
NID	g1029357
VERSION	Z58126.1 GI:1029357
KEYWORDS	Cpg island; genomic MseI fragment.
SOURCE	human.
ORGANISM	Homo sapiens

ORGANISM	Homo sapiens
Eukaryota: Metazoa: Chordata: Vertebrata: Mammalia: Eutheria: Primates: Catarrhini: Hominae: Homo.	
REFERENCE	1 (bases 1 to 291)
AUTHORS	Macdonald, M., Huckle, E., Wilkinson, P. and Micklem, G.
TITLE	Direct Submission
JOURNAL	Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire, CB10 1RO, England. E-mail contact: humquerry@sanger.ac.uk
REFERENCE	2 (bases 1 to 291)
AUTHORS	Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.
TITLE	Purification of CpG islands using a methylated DNA binding column
JOURNAL	Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE	94282070
COMMENT	vector: pGEM-5zf(-)

**FEATURES**

Resources are available from the UK MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridgeshire CB10 1HQ, UK. see URL: <http://www.hgmp.mrc.ac.uk/> for details or contact: [blomelph@hgmp.mrc.ac.uk](mailto:blomelph@hgmp.mrc.ac.uk).  
Location/Qualifiers

```

source
1. .291
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="male"
/dev_stage="adult"
/tissue_type="blood"
/clone_1b="ccl-1"
/clone="2e10"
44 a 95 c 103 g 45 t 4 others
BASE COUNT
ORIGIN

```

Query Match	41.0%	Score 282.2;	DB 9;	Length 291;
Best Local Similarity	97.6%	Pred. No. 1.7e-58;		
Matches 284;	Conservative 0;	Mismatches 7;	Indels 0;	Gaps 0;

QY	47	TAACTCGGGTTACGGGAGCCGGCTCTCGGTGGAGTCCGTTAAAGGAGCA	106
Db	1	TAAACTGGGGTTACGGGAGCCGGCTCTCTGCGGTGAGTCCGTTAAAGGAGCA	60
QY	107	GCCCGCAGGCGGCACATAGTCTCCGCACAAAGTCTCGTGCCCTTGCCATTTCCAG	166
Db	61	GNCCCGCAGGCGGCACATAGTCTCCGCACAAAGTCTCGGTGCCCCCTTGCCATTTCCAGC	120
QY	167	CGGCTCCACAGAGGGTTCACGGCGCGGGGAGAGTGTGAACCCGAGAGCTCGGCGGG	226
Db	121	CGGCTCCACAGAGGGTTCACGGCGCGGGGAGAGTGTGAACCCGAGAGCTCGGCGGG	180
QY	227	GCCCCGCTGGTGGCCCCCGGCATGAGAGGGGCTCGGGAGTGGCTCTTTTCGGCGCCC	286
Db	181	GNCCCGCTGGTGGTCCGCGGCATGAGAGGGGCTCGGGAGAGGCTCTTTTCGGCGCCC	240

QY	240	GCGGCGGCATAGACAGGGGGCTGGGACTGGGTCTTTTCCGGGGCCCTCCGGCGAGGT	299
Db	30413	GTTCGGGCCGTCACAGGGGCTCCGACGGGCTCACCTT-CGCGCCCCACCTGGCAGAAGT	30471
QY	300	GAGGGAGAGATGTCTCATGTCAAGGTTCAAGGCCAAACGAAGTACTGGCCTATCTTC	359
Db	30472	GAGATATAAATGTCTCGTGTAAGGTTCAAGGCCAAAGCTGAAGTTGTGGACTCTATCTTC	30531
QY	360	CAGGAGAACAGGAGCCACAGCGCGGCTCACGCCCCACCACCAATTAAATTA	414
Db	30532	CACAAGAACAGGAGCGCGCTGCCAGCTCATGTCTCACTCAACATCAGCTTGA	30586
RESULT	6		
AC006328		AC006328	178307 bp DNA HTG 13-MAR-1999
DEFINITION		Homo sapiens clone NH0102005, WORKING DRAFT SEQUENCE, 3 unordered pieces.	
ACCESSION		AC006328	
NID		54415969	
VERSION		AC006328.2	GI:4415969
KEYWORDS		HTG; HTGS_PHASE1.	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
AUTHORS		Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
JOURNAL		1 (bases 1 to 178307)	
REFERENCE		The sequence of Homo sapiens clone	
AUTHORS		Unpublished	
TITLE		2 (bases 1 to 178307)	
JOURNAL		Waterston,R.H.	
REFERENCE		Submitted (09-JAN-1999) Genome Sequencing Center, Washington	
AUTHORS		University School of Medicine, 4444 Forest Park Parkway, St. Louis,	
JOURNAL		MO 63108, USA	
COMMENT		On Mar 13, 1999 this sequence version replaced gi:4138780. * NOTE: This is a 'working draft' sequence. It currently * consists of 3 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 1163: contig of 1163 bp in length * 11164 1181: gap of unknown length * 11182 5706: contig of 4586 bp in length * 57067 57084: gap of unknown length * 57085 178307: contig of 12123 bp in length. Location/Qualifiers 1. 178307 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="NH0102005"	
BASE COUNT		49986 a 38822 c 39355 g 50108 t 36 others	
ORIGIN			
FEATURES			
source			
Query Match		37.9%; Score 261; DB 35; Length 178307;	
Best Local Similarity		80.0%; Pred. No. 2,7e-53;	
Matches 332; Conservative 0; Mismatches 80; Indels 3; Gaps 2;			
QY	1	AGCCCTTCCAGGCGGTAGCGGGGCGAGTGTGCTGTGCCCTTTAACTGCGGCTTG	60
Db	10173	AGCCCTTCCAGGCGGTAGCGGGGCGTGTGTCTGTGCTTAACTGCGGCTTG	10232
QY	61	ACGGGACCGCGGCTCTGTGCGTGTGAGTGTGCTATTAAGGAGACAGCGCGGCGGC	120
Db	10233	ACAGGACGAGGCTCTGTGCGTGTGAGTGTGCTATTAAGGAGGAGAGCGCGCGGCTC	10292
QY	121	CACATACCT-CCGCGCAAAGTCTTGATGCCCTTGCATTTTCCAGCGCGCTCCACGA	179

```

Db 10293 CACACAGCTCCCCGAGAGGCGCTTGTCCTTGCATTTTCAGACCCCTACTGTGACTA 10352
Qy 180 GGGTCACGGCGGCGGAGAGGTGAGCCCGAGAGCTTCGCCGGGGGGCCCCCGCTGTG 239
Db 10353 GAGTTGAGCAGCAGAGAGGAGCGTGGAGAGCAGACCCGAGAGAGTCCCGCGGGTG 10412
Qy 240 GCGCGGGCCATGACAGCGGCTGGGAGCTGGCTCTTTCCGCCGCCCTCCCGCGAGGT 299
Db 10413 GTTGGCGCCGTACAGCGGCTCCG--ACGGGCTCACCTTCGCCGCCACCTCCGAGGT 10470
Qy 300 GAGGAGGAGATGTCATGTCAGGTTTCAAGGCCCAACCAAGTACTGCGCTATCTTC 359
Db 10471 GAGATTAATATCTCGTGTGAGGTTTCAAGGCCCAAGTACTGTCAGTCTATCTTC 10530
Qy 360 CAGGAGAACGAGAGCAGAGCGCGGCTCACGCCCAACCCGACATTAAGATTA 414
Db 10531 CACAAGAACGAGAGCGCTGCCGAGCTCATGTCTCACTGCACATCAGCTTGA 10585

LOCUS HS2G7F 298 bp DNA PRI 19-OCT-1995
DEFINITION H.sapiens Cpg island DNA genomic MseI fragment, clone 2g7, forward
ACCESSION Z60646
VERSION 260646.1 GI:1032750
KEYWORDS Cpg island; genomic MseI fragment.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 298)
AUTHORS MacDonald,M., Huckle,E., Wilkinson,P. and Mcklem,G.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk
2 (bases 1 to 298)
AUTHORS Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
TITLE Purification of Cpg islands using a methylated DNA binding column
JOURNAL Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE 94282070
COMMENT Vector: pGEM-52f(-)
Clones are available from the UK MRC Human Genome Mapping Project
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
http://www.hgmp.mrc.ac.uk/ for details
or contact: biolhelp@hmp.mrc.ac.uk.
FEATURES
source location/Qualifiers
1..298
/db_xref="taxon:9606"
/organism="Homo sapiens"
/sex="male"
/dev_stage="adult"
/tissue_type="blood"
/clone_lib="CGI-1"
/clone="2g7"
BASE COUNT 47 a 99 c 102 g 48 t 2 others
ORIGIN
Query Match 34.4%; Score 237.2; DB 9; Length 298;
Best Local Similarity 94.7%; Pred. No. 1.2e-47;
Matches 288; Conservative 0; Mismatches 10; Indels 6; Gaps 4;
Qy 48 AACGCGGCTGAGGAGCGCGGCTTCGCTGAGGATGATTAAGGAGAGAG 107
Db 1 AATGCGGCTGAGGAGCGCGGCTTCGCTGAGGATGATTAAGGAGAGAG 59
Qy 108 CCGCGGAGGCGGAGAGTCCGCGGAGTCTGCTGCTTGCATTTTCAGGC 167
Db 60 CCGCGGA-GCGCGCAGATAGCTCCGCGAGTCTCGGCTTGCATTTTCAGGC 118
Qy 168 GCGCTCCAGAGGTACGCGCGGAGAGGTGAGACCGGAGAGCTCGCGGGG 227

```

```

Db 119 GCGTCCACGAGGCTACGCGGCGGAGAGGTGGAGCCGCGAGANTCGCC--GGG 175
Qy 228 CCGCGCTGTGGCGCGGCGGAGAGGCTGAGAGGCTGCTCTTTTCCGCCCT 287
Db 176 GCGCGCTGTGGCGCGGCGGAGAGGCTGAGAGGCTGCTCTTTTCCGCCCT 235
Qy 288 CCGCGGAGGTGAGGAGAGATGTCATGTCAGGTTCAAGGCCAAACGAGTACTG 347
Db 236 CCGCGCGGA-GTGAAGGAGAGATGTCATGTCAGGTTCAAGGCCAAACGAGTACTG 294
Qy 348 GCGT 351
Db 295 GCGT 298

RESULT 8
LOCUS AF036903 1712 bp mRNA PRI 16-OCT-1998
DEFINITION Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor
ACCESSION AF036903
VERSION 93757807
KEYWORDS AF036903.1 GI:3757807
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1712)
AUTHORS Gault,J., Robinson,M., Berger,R., Drebing,C., Logel,J., Hopkins,J.,
Moore,T., Jacobs,S., Meliether,J., Choi,M.J., Kim,E.J., Walton,K.,
Bulling,K., Davis,A., Breese,C., Freedman,R. and Leonard,S.
TITLE Genomic organization and partial duplication of the human alpha7
neural nicotinic acetylcholine receptor gene (CHRNA7)
JOURNAL Genomics 52 (2), 173-185 (1998)
MEDLINE 99000837
REFERENCE 2 (bases 1 to 1712)
AUTHORS Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R.,
Breese,C., Davis,A., Hopkins,J. and Freedman,R.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1997) Psychiatry, University of Colorado Health
Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
FEATURES
source location/Qualifiers
1..1712
/db_xref="taxon:9606"
/organism="Homo sapiens"
/db_xref="dbEST:M52861"
/db_xref="dbEST:AA037389"
/chromosome="15"
/clone="321242"
/clone_lib="Soares papathroid tumor NbHPA library,
Research Genetics/IMAGE Consortium, LLNL"
<1..1712
/note="alpha-7 neuronal nicotinic acetylcholine receptor;
alternately spliced"
BASE COUNT 356 a 489 c 457 g 410 t
ORIGIN
Query Match 22.6%; Score 155.4; DB 11; Length 1712;
Best Local Similarity 99.4%; Pred. No. 6.3e-28;
Matches 156; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 533 CAATTCATATGCTAATTCACATTTTGGATCTGCAAACTCGATATGCTGATGAG 592
Db 45 CAGTTCATATGCTAATTCACATTTTGGATCTGCAAACTCGATATGCTGATGAG 104
Qy 593 GCGTTGAGCGGAGATTCACACACTAAGTGTGATGATTTCTGGGCAATGCGAGTAC 652
Db 105 GCGTTGAGCGGAGATTCACACACTAAGTGTGATGATTTCTGGGCAATGCGAGTAC 164
Qy 653 CTGCTCCAGGAGATTCACAGAGTTCCTGCTACATCG 689

```

Db	165	CTGCCTCCAGGCAATTCACAGATTCTCGTCAATCG	201
RESULT	9		
AF037646			
LOCUS			
DEFINITION	Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor precursor RNA, partial sequence.	PRI	16-OCT-1998
ACCESSION	AF037646		
NID	93757808		
VERSION	AF037646.1	GI:3757808	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
AUTHORS	Gault,J., Robinson,M., Berger,R., Drebing,C., Logel,J., Hopkins,J., Moore,T., Jacobs,S., Merlither,J., Choi,M.J., Kim,E.J., Walton,K., Bulting,K., Davis,A., Breese,C., Freedman,R. and Leonard,S. Genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene (CHRNA7)		
JOURNAL	Genomics 52 (2), 173-185 (1998)		
MEDLINE	99000837		
FEATURES	2 (bases 1 to 1836) Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R. Breese,C., Davis,A., Hopkins,J. and Freedman,R. Direct Submission Submitted (03-DEC-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA Location/Qualifiers 1..1836 /organism="Homo sapiens" /db_xref="taxon:9606" /db_xref="dbEST:W03952" /db_xref="dbEST:N73891" /chromosome="15" /clone="297134" /clone.lib="Soares fetal liver spleen library, Research Genetics/IMAGE Consortium, LNL" /note="unprocessed mRNA with intron" <1..1836 /note="alpha-7 neuronal nicotinic acetylcholine receptor precursor; intron included in 5' region"		
TITLE	Submitted (03-DEC-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA		
JOURNAL			
FEATURES			
SOURCE			
misc_feature			
BASE COUNT	383 a 503 c 469 g 481 t		
ORIGIN			
very Match	22.4%; Score 154.4; DB 11; Length 1836;		
Best Local Similarity	99.4%; Pred. No. 1.1e-27;		
Matches 155; Conservative	0; Mismatches 1; Indels 0; Gaps		
QY	534 AATTCATTGTCTAATTCACAGATTGTGGATAGCTGCAAACTGGATATTGTGTAGAGC	593	
DB	168 AGTTCAATTGTGTATCATTCAGCATTTGTGGATAGCTGCCAACTGGATATTGTGTAGAGC	227	
QY	594 GCTTGAGCCACATTCACACTAACGGTTGGTAGATTCTTCGGCATTCGGCATTCAGTACC	653	
DB	228 GCTTGAGCCACATTCACACTAACGGTTGGTAGATTCTTCGGCATTCGGCATTCAGTACC	287	
QY	654 TGCCCTCAGGCATATTCAGAAGTCTGCTCAATCG	689	
DB	288 TGCCTCAGGCATATTCAGAAGTCTGCTCAATCG	323	
RESULT	10		
HS11562F			
LOCUS	HS11562F 180 bp DNA PRI 17-OCT-1995		
DEFINITION	H.sapiens Cpg island DNA genomic MseI fragment, clone 11562,		
ACCESSION	Z54415		
NTD	g1020456		

VERSION	Z54415.1 GI:1020456									
KEYWORDS	CpG island; genomic MseI fragment.									
SOURCE	human.									
ORGANISM	Homo sapiens									
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;									
AUTHORS	Primates; Catarrhini; Hominoidea; Homo.									
TITLE	1 (bases 1 to 180)									
JOURNAL	Macdonald,M., Huckle,E., Wilkinson,P. and Micklem,G.									
REFERENCE	Direct Submission									
AUTHORS	Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,									
TITLE	CB10 IRQ, England. E-mail contact: humquerry@sanger.ac.uk									
JOURNAL	2 (bases 1 to 180)									
REFERENCE	Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.									
AUTHORS	Purification of CpG islands using a methylated DNA binding column									
TITLE	Nat. Genet. 6 (3), 236-244 (1994)									
JOURNAL	94282070									
MEDLINE	Vector: pGEM-5zf(-)									
COMMENT	Clones are available from the UK MRC Human Genome Mapping Project									
	Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:									
	http://www.hgmp.mrc.ac.uk/for details									
	or contact: biohelp@hgmp.mrc.ac.uk.									
FEATURES	location/Qualifiers									
SOURCE	1..180									
	/organism="Homo sapiens"									
	/db_xref="taxon:9606"									
	/sex="male"									
	/dev_stage="adult"									
	/tissue_type="blood"									
	/clone_id="CGI-1"									
	/clone="11592"									
BASE COUNT	33 a 57 c 58 g 29 t 3 others									
ORIGIN										
Query Match	17.2%; Score 118.8; DB 9; Length 180;									
Best Local Similarity	87.0%; Pred. No. 3.8e-19;									
Matches 140; Conservative	0; Mismatches 20; Indels 1; Gaps 1;									
QY	47 TAACTGGCGCTTACGGGAGCGCGCTCTCTGCGGAGTGGTTATTAAGGAGCA 106									
Db	1 TAACTGGCGTTTACAGGGGCCGCTCTCTGCGGAGTGGTTACAAAGGAGCA 60									
QY	107 GCCCGAGGCGCGCACATAGCTCCGCGCAAGTCTCTGCGTCCCTTGCATTTCAGC 166									
Db	61 GCCCGAGGCGCGCACAGACTCCCGCAAGACT-CTGCTCCCTTGCATTTCAGC 119									
QY	167 CGCGCTCCACGAGGCTACGCGCGCGGAGGAGGTGAGC 207									
Db	120 CGCGCTCCACGAGAGTGGAGGCTGCAGAGAGCGGNTC 160									
RESULT 11										
LOCUS	HS165H1F									
DEFINITION	H.sapiens CpG island DNA genomic MseI fragment, clone 165h1,									
ACCESSION	257156									
NID	91028387									
VERSION	257156.1 GI:1028387									
KEYWORDS	CpG island; genomic MseI fragment.									
SOURCE	human.									
ORGANISM	Homo sapiens									
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;									
AUTHORS	Primates; Catarrhini; Hominoidea; Homo.									
TITLE	1 (bases 1 to 171)									
JOURNAL	Dodsworth,S.J., Huckle,E., Wilkinson,P. and Micklem,G.									
REFERENCE	Direct Submission									
AUTHORS	Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,									
TITLE	CB10 IRQ, England. E-mail contact: humquerry@sanger.ac.uk									
JOURNAL	2 (bases 1 to 171)									
REFERENCE	Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.									
AUTHORS	Purification of CpG islands using a methylated DNA binding column									
TITLE	Nat. Genet. 6 (3), 236-244 (1994)									
JOURNAL	94282070									
MEDLINE	Vector: pGEM-5zf(-)									
COMMENT	Clones are available from the UK MRC Human Genome Mapping Project									
	Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:									
	http://www.hgmp.mrc.ac.uk/for details									
	or contact: biohelp@hgmp.mrc.ac.uk.									
FEATURES	location/Qualifiers									
SOURCE	1..180									
	/organism="Homo sapiens"									
	/db_xref="taxon:9606"									
	/sex="male"									
	/dev_stage="adult"									
	/tissue_type="blood"									
	/clone_id="CGI-1"									
	/clone="11592"									
BASE COUNT	33 a 57 c 58 g 29 t 3 others									
ORIGIN										



Db 445 CTATACAGTCTGATGAGCCCTTGACGCACATTCACACTACAGTGTGTAATTC 504  
OY 634 TTTCGGGATGGCCAGTACCTGCCTCCAGCATATTCAGAGTCCGTGCATCG 689  
Db 505 TTCTGGGATGGCCAGTACCTGCCTCCAGCATATTCAGAGTGTGTAATTC 560

RESULT 14  
LOCUS HSNACHRA7 1509 bp mRNA PRI 22-JAN-1998  
DEFINITION H.sapiens mRNA for nicotinic acetylcholine receptor alpha7 subunit precursor.  
ACCESSION Y08420  
NID 92808623  
VERSION Y08420.1 GI:2808623  
KEYWORDS nAChR gene; nicotinic acetylcholine receptor alpha 7 subunit.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Euthera;  
Primates; Catarrhini; Homindae; Homo.  
1 (bases 1 to 1509)  
Groot Kormelink, P.J. and Luyten, W.H.  
Cloning and sequence of full-length cDNAs encoding the human neuronal nicotinic acetylcholine receptor (nAChR) subunits betas3 and beta4 and expression of seven nAChR subunits in the human neuroblastoma cell line SH-SY5Y and/or IMR-32  
FEBS Lett. 400 (3), 309-314 (1997)  
JOURNAL 9716223  
MEDLINE 2 (bases 1 to 1509)  
REFERENCE Groot Kormelink, P.J.  
AUTHORS Direct Submission  
TITLE Submitted (27-SEP-1996) P.J. Groot Kormelink, Janssen Research Foundation, Exp. Mol. Biol. Dept., Turnhoutseweg 30, B-2340 Beere, Belgium  
REMARK Revisited by author 22-JAN-1998  
COMMENT On Jan 25, 1998 this sequence version replaced gi:1702915.  
FEATURES  
source 1..1509  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="neuroblastoma"  
/cell\_line="IMR-32"  
/cell\_line="SH-SY5Y"  
1..66  
/gene="nAChRA7"  
1..1509  
/gene="nAChRA7"  
/codon\_start=1  
/product="nicotinic acetylcholine receptor alpha7 subunit precursor"  
/protein\_id="CAA69697.1"  
/db\_xref="PID:e1246208"  
/db\_xref="PID:g2808624"  
/db\_xref="GI:2808624"  
/translation="MRCSPGGVWLAALASLHVSLOGEFORLKYELKYNPNLEERV ANDQPLTVFSLIQIMVDENKQVLTINIMLOMSMTDHYLOMNSVPGKYVFPDGOIMKPDILYNSADERFADATFHTNVLVNSGHCQYLPGRFSKSCYIDVWMPED VOHCKLFGSMGSLDQMOEADISGIPNGEMDNLGIPGRSREREYECCKEPP DVITVTYMRRTLYGSLNLIPCVLISALALVFLPADSGEKISGITVLSTIVFM LVAELMPATSDSVPLIAQYFASMTIIVGLSVVTVIVLOYHHHDPGGKMPKWTYV LLMNCAMFLRMKRGEDKVRPAQHKORCSLASVMSAVAPPASNGNLVIGFRGL DGVCVPTPDGSGVYCGRMACSPTHDEHLHGQPPEDPLAKILEEVRYIARFRCO DESAVCSSEMKFACVYDRLCLMAFSVFTICTIGILMSAPNFEAVSNDFA"  
gene 1..1509  
/gene="nAChRA7"  
mat\_peptide 67..1506  
/gene="nAChRA7"  
BASE COUNT 297 a 451 c 429 g 332 t  
ORIGIN  
Query Match 15.7%, Score 108, DB 10, Length 1509;

Best Local Similarity 95.7%; Pred. No. 1.6e-16;  
Matches 111; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 574 CTGCGATATGCGATGAGGCGTTTGAGCCACATTCACAGTGTGTAATTC 633  
Db 342 CTATACAGTCTGATGAGGCGTTTGAGCCACATTCACAGTGTGTAATTC 401

OY 634 TTTCGGGATGGCCAGTACCTGCCTCCAGCATATTCAGAGTCCGTGCATCG 689  
Db 402 TTCTGGGATGGCCAGTACCTGCCTCCAGCATATTCAGAGTGTGTAATTC 457

RESULT 15  
LOCUS HS062436 1876 bp mRNA PRI 11-JAN-1997  
DEFINITION Human nicotinic acetylcholine receptor alpha7 subunit precursor, mRNA, complete cds.  
ACCESSION U62436  
NID 91458119  
VERSION U62436.1 GI:1458119  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Euthera; Primates; Catarrhini; Homindae; Homo.  
1 (bases 1 to 1876)  
Elliot, K.J., Ellis, S.B., Berckhan, K.J., Urrutia, A., Chavez-Noriega, L.E., Johnson, E.C., Velicelebi, G. and Harpold, M.M.  
Comparative structure of human neuronal alpha 2-alpha 7 and beta 2-beta 4 nicotinic acetylcholine receptor subunits and functional expression of the alpha 2, alpha 3, alpha 4, alpha 7, beta 2, and beta 4 subunits  
J. Mol. Neurosci. 7 (3), 217-228 (1996)  
JOURNAL 97062879  
MEDLINE 2 (bases 1 to 1876)  
REFERENCE Elliot, K.J.  
AUTHORS Direct Submission  
TITLE Submitted (28-JUN-1996) Kathryn J. Elliot, SIBRA Neurosciences, Inc., 505 Coast Blvd. So., La Jolla, CA 92037, USA  
FEATURES  
source 1..1876  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="SIBRA lambda gt11 library #3 (M. Williams)"  
/cell\_line="IMR32"  
/cell\_type="neuroblastoma"  
1..72  
/gene="nAChRA7"  
73..138  
73..1581  
/codon\_start=1  
/product="nicotinic acetylcholine receptor alpha7 subunit precursor"  
/protein\_id="AAB40114.1"  
/db\_xref="PID:g1458120"  
/db\_xref="GI:1458120"  
/translation="MRCSPGGVWLAALASLHVSLOGEFORLKYELKYNPNLEERV ANDQPLTVFSLIQIMVDENKQVLTINIMLOMSMTDHYLOMNSVPGKYVFPDGOIMKPDILYNSADERFADATFHTNVLVNSGHCQYLPGRFSKSCYIDVWMPED VOHCKLFGSMGSLDQMOEADISGIPNGEMDNLGIPGRSREREYECCKEPP DVITVTYMRRTLYGSLNLIPCVLISALALVFLPADSGEKISGITVLSTIVFM LVAELMPATSDSVPLIAQYFASMTIIVGLSVVTVIVLOYHHHDPGGKMPKWTYV LLMNCAMFLRMKRGEDKVRPAQHKORCSLASVMSAVAPPASNGNLVIGFRGL DGVCVPTPDGSGVYCGRMACSPTHDEHLHGQPPEDPLAKILEEVRYIARFRCO DESAVCSSEMKFACVYDRLCLMAFSVFTICTIGILMSAPNFEAVSNDFA"  
mat\_peptide 139..1578  
/product="nicotinic acetylcholine receptor alpha7 subunit"  
BASE COUNT 369 a 553 c 531 g 423 t  
ORIGIN  
Query Match 15.7%, Score 108, DB 10, Length 1876;

Best Local Similarity 95.7%; Pred. No. 1.6e-16;  
Matches 111; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 574 CTGGGATATTGCTGATGAGCGCTTTGACGCCACATTCCACACTAACGTGTGGTGAATTC 633

Db 414 CTATAACAGTGGCTGATGAGCGGCTTGAGCGCCACATTCCACACTAACGTTGGTGATTC 473

QY 634 TTCTGGGCATTGCCAGTACCTGCTCCAGGCATATTCAAGAGTTCCTGCTACATCG 689

Db 474 TTCTGGGCAITGGCAGTACCTGCGCTCCAGGCATATTCAAGAGTTCCCTGCTACATCG 529

Search completed: September 17, 1999, 22:05:19  
Job time: 16502 sec





PI Elliott KJ, Ellis SB, Harpold MM;  
 DR WPI: 94-303024/37.  
 P-PsDB: W44153.  
 CC Human neuronal nicotinic acetylcholine receptor subunits and DNA -  
 PT also transformed cells useful for screening cpts. which modulate  
 activity of the receptor  
 PS Claim 8: Page 78-79; 99pp; English.  
 CC The present sequence encodes a human neuronal nicotinic acetylcholine  
 CC receptor (NACHR) subunit. The cells expressing the alpha and/or beta  
 CC NACHR subunits may be used in a method of screening compounds to  
 CC identify any which modulate the activity of human neuronal NACHR.  
 CC Subunit specific antibodies may be used to monitor the distribution  
 CC and expression density of various subunits in normal vs diseased brain  
 CC tissues. Testing of single receptor subunits in normal vs diseased brain  
 CC subunit combinations with a variety of potential agonists or antagonists  
 CC provides information with respect to the function and activity of the  
 CC individual subunits and should lead to the identification and design of  
 CC compounds that are capable of very specific interaction with one or  
 CC more receptor subtypes. The resulting drugs should exhibit fewer  
 CC unwanted side effects than drugs identified e.g. screening with cells  
 CC that express a variety of subtypes.  
 CC Sequence 1876 BP; 369 A; 553 C; 530 G; 423 T;  
 Query Match 15.7%; Score 108; DB 1; Length 1876;  
 Best Local Similarity 95.7%; Pred. No. 4.8e-21;  
 Matches 111; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 OY 574 CTGCATATTGCTGATGAGCGCTTTGAGCCCATTCACACTACGATGTTGGTGAATTC 633  
 DB 414 CTATACAGCTGCTGATGAGCGCTTTGAGCCCATTCACACTACGATGTTGGTGAATTC 473  
 OY 634 TTCTGGGCAATGGCCAGTACCTGCTCCAGGCAATTCAGAGTCTCTCTACATCG 689  
 DB 474 TTCTGGGCAATGGCCAGTACCTGCTCCAGGCAATTCAGAGTCTCTCTACATCG 529  
 RESULT 3  
 V44687 Location/Qualifiers  
 ID V44687 standard; cDNA; 1590 BP.  
 AC V44687;  
 DT 09-OCR-1998 (first entry)  
 DE V274T variant human alpha7 nACHR coding sequence.  
 KW Alpha7 nACHR; alpha7 nicotinic acetylcholine receptor subunit; cancer;  
 KW neurodegeneration; enzyme dysfunction; affective disorder; therapy;  
 KW immune dysfunction; diabetic neuropathy; Alzheimer's disease;  
 KW schizophrenia; ss.  
 OS Homo sapiens.  
 Key  
 CDS 9..1517  
 F1 /\*tag- a  
 F1 WO9828331-A2.  
 F1 02-JUL-1998.  
 PF 22-DEC-1997; U23405.  
 PR 20-DEC-1996; US-771737.  
 PA (ABBO ) ABBOTT LAB.  
 PI Briggs CA, Gopalakrishnan M, McKenna DG, Monteggia LM,  
 PI Roch J, Sullivan JP, Touma E;  
 DR WPI: 98-377593/32.  
 DR P-PsDB: W69216.  
 CC Nucleic acid encoding variant of human alpha7 nicotinic  
 CC acetylcholine receptor sub-unit - used to identify modulators of  
 CC the receptor, potentially useful for treating neuro-degeneration,  
 CC cancer, affective disorders etc.  
 PS Claim 14: Fig 2; 44pp; English.  
 CC This sequence encodes the V247T variant of human alpha7 nicotinic  
 CC acetylcholine receptor (NACHR) subunit of the invention. Cells containing  
 CC the DNA are used to express the protein and to identify modulators of  
 CC alpha7 nACHR activity or cytoprotective agents, e.g. antisense  
 CC compounds or antagonists that are potentially useful for treating  
 CC neurodegeneration, enzyme dysfunction, affective disorders and immune  
 CC dysfunction, such as cancer, post-herpetic neuralgia, diabetic  
 CC neuropathy, osteoarthritis, Alzheimer's or Parkinson's diseases, kuru,

CC psychosis and schizophrenia. Probes based on the DNA are used to detect  
 CC the DNA in usual hybridization or amplification tests, while monoclonal  
 CC antibodies are used to detect the protein for diagnosis (in vitro or by  
 CC in situ immuno-fluorescent assay). Compared with wild-type alpha7 nACHR,  
 CC the protein has about 100-fold greater sensitivity to cholinergic  
 CC receptor agonists (nicotine or acetylcholine) and response to these  
 CC agonists decays more slowly, but the wild-type inward rectification is  
 CC retained.  
 CC Sequence 1590 BP; 322 A; 471 C; 449 G; 348 T;  
 Query Match 15.7%; Score 108; DB 1; Length 1590;  
 Best Local Similarity 95.7%; Pred. No. 4.5e-21;  
 Matches 111; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 OY 574 CTGCATATTGCTGATGAGCGCTTTGAGCCCATTCACACTACGATGTTGGTGAATTC 633  
 DB 350 CTATACAGCTGCTGATGAGCGCTTTGAGCCCATTCACACTACGATGTTGGTGAATTC 409  
 OY 634 TTCTGGGCAATGGCCAGTACCTGCTCCAGGCAATTCAGAGTCTCTCTACATCG 689  
 DB 410 TTCTGGGCAATGGCCAGTACCTGCTCCAGGCAATTCAGAGTCTCTCTACATCG 465  
 RESULT 4  
 ID T59196 standard; cDNA; 2769 BP.  
 AC T59196;  
 DT 17-JUN-1997 (first entry)  
 DE Neuronal alpha-bungarotoxin binding protein alpha subunit cDNA.  
 KW Neuronal alpha-bungarotoxin binding protein alpha 1; cholinergic;  
 KW ligand binding; ion channel; ss.  
 OS Gallus sp.  
 FH Key  
 FT cgs Location/Qualifiers  
 FT 71..1513  
 FT /\*tag- a  
 FT signal\_peptide 71..136  
 FT /\*tag- b  
 FT mat\_peptide 137..1510  
 FT /\*tag- c  
 PN US5599709-A.  
 PD 04-FEB-1997.  
 PF 28-SEP-1989; 413947.  
 PR 28-SEP-1989; US-413947.  
 PI (SALK ) SALK INST BIOLOGICAL STUDIES.  
 PI Lindstrom JM, Schoepfer RD;  
 DR WPI: 97-118297/11.  
 DR P-PsDB: W12368.  
 CC New isolated neuronal alpha-bungarotoxin-binding protein DNA - used  
 CC to screen cholinergic agents and other drugs which may affect ligand  
 CC binding, ion channel or other activities of the protein.  
 PS Claim 1; Fig 2A-B; 18pp; English.  
 CC 2 cDNA clones (T59196 and T59197), respectively code for the alpha1  
 CC subunit (W12368) and alpha2 subunit (W12369) of chick neuronal  
 CC alpha-bungarotoxin binding protein (ABBP). They were isolated from  
 CC an 18-day embryo chick brain cDNA library using a probe (see also  
 CC T59198) based on the N-terminal amino acid sequence of chicken  
 CC brain ABBP. The probe isolated partial clone pch29-1, which  
 CC encoded the N-terminal portion of alpha1. A subclone, pch29-3  
 CC (ATCC 40641), was used to rescreen the library, yielding clone  
 CC pch31-1 (ATCC 40640), which encoded the entire alpha2 sequence.  
 CC A probe based on the C-terminal region of pch31-1 was used obtain  
 CC clone pch34-1 (ATCC 40639), encoding the C-terminal portion of  
 CC alpha1. The cDNA clones can be used as probes to identify further  
 CC ABBP subunits, and in the recombinant prodn. of ABBP.  
 CC Sequence 2769 BP; 750 A; 604 C; 571 G; 804 T;  
 Query Match 10.8%; Score 74.4; DB 1; Length 2769;  
 Best Local Similarity 77.6%; Pred. No. 1.5e-11;  
 Matches 90; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
 OY 574 CTGCATATTGCTGATGAGCGCTTTGAGCCCATTCACACTACGATGTTGGTGAATTC 633

DB 412 CTATACAGTCTGATGATTAAGATTGCTACATTTCACACTAATGTTTACATATTC 471  
 QY 634 TTCTGGGACATTCGACGATTCGCTCCAGGACATTCACAGTTCCTCCGATACG 689  
 DB 472 TTGGGACACTGCGCATATCTGCGACGACGACGATTTAAAGCTACGTCATAG 527

## RESULT 5

ID T59197 standard; cDNA; 2101 BP.  
 AC T59197;  
 DE 17-JUN-1997 (first entry)  
 DE Neuronal alpha-bungarotoxin binding protein alpha2 subunit cDNA.  
 KW Neuronal alpha-bungarotoxin binding protein alpha 2; cholinergic;  
 OS Gallus sp.  
 Key Location/Qualifiers  
 cds 56..1501  
 signal\_peptide 56..145  
 mat\_peptide 146..1498  
 /\*tag= a  
 /\*tag= b  
 /\*tag= c

FT signal\_peptide 56..145  
 FT mat\_peptide 146..1498  
 FT /\*tag= a  
 FT /\*tag= b  
 FT /\*tag= c  
 PD US559709-A.  
 PD 04-FEB-1997.  
 PF 28-SEP-1989; 413947.  
 PR 28-SEP-1989; US-413947.  
 PA (SALK ) SALK INST BIOLOGICAL STUDIES.  
 PI Lindstrom JM, Schoepfer RD;  
 DR WPI: 97-118297/11.  
 DR P-PDB: W12369.  
 PT New isolated neuronal alpha-bungarotoxin-binding protein DNA - used  
 PT binding, ion channel or other activities of the protein.  
 PS Claim 1; Fig 3A-B; 18pp; English.  
 CC 2 cDNA clones (T59196 and T59197) respectively code for the alpha1  
 CC subunit (W12368) and alpha2 subunit (W12369) of chick neuronal  
 CC alpha-bungarotoxin binding protein (ABBP). They were isolated from  
 CC an 18-day embryo chick brain cDNA library using a probe (see also  
 CC T59198) based on the N-terminal amino acid sequence of chicken  
 CC brain ABBP. The probe isolated partial clone pCh29-1, which  
 CC encoded the N-terminal portion of alpha1. A subclone, pCh29-3  
 CC (ATCC 40641), was used to rescreen the library, yielding clone  
 CC pCh31-1 (ATCC 40640), which encoded the entire alpha2 sequence.  
 CC A probe based on the C-terminal region of pCh31-1 was used obtain  
 CC clone pCh34-1 (ATCC 40639), encoding the C-terminal portion of  
 CC alpha1. The cDNA clones can be used as probes to identify further  
 CC ABBP subunits, and in the recombinant prodn. of ABBP.  
 SQ Sequence 2101 BP; 582 A; 406 C; 454 G; 659 T;

Query Match  
 Best Local Similarity 8.9%; Score 61.6; DB 1; Length 2101;  
 Matches 82; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 574 CTGGCATATTCGATGAGGCGTTTGACGCGACATTCACACTAAGTGTGTCATTC 633  
 DB 421 CTATACAGTCTGATGATTAAGATTGCTACATTTCACACTAAGTGTGTCATTA 480  
 QY 634 TTCTGGGACATTCGACGATTCGCTCCAGGACATTCACAGTTCCTCCGATACG 689  
 DB 481 CTCCTGATCTCTGCATATATTCCTCCAGGACATTTGAAGACACATGTTCACTG 536

## RESULT 6

V68408/c  
 ID V68408 standard; cDNA to mRNA; 5561 BP.  
 AC V68408;  
 DE 05-MAY-1999 (first entry)  
 DE Human BAZ1-beta cDNA #1.  
 KW Transcriptional regulator; BAZ1-alpha; bromodomain; BAZ;  
 KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;  
 PI Jones MH;

KW BAZ2-beta; treatment; cancer; proliferative disorder; screening; ds.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 346..4929  
 FT /\*tag= a  
 FT /\*product= "BAZ1-beta"

PN W09847920-A1.  
 PD 29-OCT-1998.  
 PF 17-APR-1998; J01783.  
 PR 24-OCT-1997; JP-310027.  
 PR 18-APR-1997; JP-116570.  
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
 PI Jones MH;  
 DR WPI: 98-583603/49.  
 DR P-PDB: W81172.  
 PT Transcriptional regulator gene family containing bromodomain - may  
 PT be expressed in testis tissue and is useful for treatment of cancer  
 PT and other proliferative disorders  
 PS Claim 2; Page 125-137; 187pp; Japanese.  
 CC This sequence encodes the human BAZ1-beta protein, a member of a  
 CC family of transcriptional regulator genes containing a bromodomain (BAZ,  
 CC in testis tissue and also in certain tumour lines. Transgenic cells may  
 CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and  
 CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer  
 CC and other proliferative disorders, and in screening of compounds for  
 CC their binding ability to the expression products (e.g. for use as drugs  
 CC by modulation of transcriptional regulation).  
 SQ Sequence 5561 BP; 1685 A; 1261 C; 1471 G; 1164 T;

Query Match  
 Best Local Similarity 6.8%; Score 46.8; DB 1; Length 5561;  
 Matches 111; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 10 CAGCGCGTAGCGGGGCGGACGCTGCTGCTTAACTGCGGCTGACGGAGCC 69  
 DB 264 CAGAGCGCGGGGGTGGGTGGGAGAGGAGGCGGCGGAGACTCCGGC 205  
 QY 70 GCCCTCTCTCGGTGATGCTGTTAAAGGAGCAGCCCGGAGCGCCACATAGCT 129  
 DB 204 TCCCTACGCCCGCGCGCGCGACAGTATGAGCGGAGAGCGGACAGCCAGACT 145  
 QY 130 CCGGCAAGTCTCGGAGGCCCTTGCATTTTCACCGCGCTCCACAGAGGTACAGC 189  
 DB 144 CCGGCGACACCGCGCGCTCCACAGACCCCGCCGACCTTCGCTGCGTCCCGC 85  
 QY 190 GGGGGGAGAGGTGAGCGCGGAGAGCTCGCGCGGGG 227  
 DB 84 GCGCGGAGTCAAGACTCTCTCTCAGACAGCCGAGG 47

## RESULT 7

V68409/c  
 ID V68409 standard; cDNA to mRNA; 5573 BP.  
 AC V68409;  
 DE 05-MAY-1999 (first entry)  
 DE Human BAZ1-beta cDNA #2.  
 KW Transcriptional regulator; BAZ1-alpha; bromodomain; BAZ;  
 KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;  
 KW BAZ2-beta; treatment; cancer; proliferative disorder; screening; ds.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 346..4941  
 FT /\*tag= a  
 FT /\*product= "BAZ1-beta"



```

FT      polya-signal      /number-3
FT      2436..2441
FT      /*tag-1
FT      conflict          replace(268..270,gac)
FT      /*tag-1
FT      /note="conflict between genomic and cDNA sequences"
PN      MO9010648-A.
PD      20-SEP-1990.
PF      14-MAR-1990; U01403.
PR      14-MAR-1989; US-321384.
PR      12-MAR-1990; US-492555.
PA      (SALK ) SALK INST FOR BIOL. STUD.
PI      Heimann SF, Denetis ES; Duvoisin RM; Patrick JW;
DR      WPI; 90-304987/40.
DR      P-PSDB; R07143.
PT      New neuronal nicotinic acetylcholine receptor - compns. contg.
PT      beta 4 sub-unit and DNA sequences encoding them.
F-      Claim 11: Fig 3; 47pp; English.
        The sequence encodes a novel neuronal nicotinic acetylcholine
        receptor subunit, beta 4. Plasmid pPC13 (ATCC 67893) was
        isolated from a cDNA library in lambda zap II prep'd. from rat
        mRNA. The sequence of the cDNA in the region encoded by the 5th
        exon is identical to that determined for the genomic clone with
        the exception of a subst. at posn. 720 of a t for a c (this does
        not alter the AA sequence); this may be due to a polymorphism
        between the rat strains used for the genomic library and the cDNA
        library, or could have resulted from a reverse transcriptase error
        during prep'n. of the cDNA. A sequence, the core of which is 44 bp
        long is repeated three times at the beginning of the 3' UT region.
        The function of this remains unknown. The protein subunit can
        combine with the known subunits, alpha-2,-3, and -4 and beta-2
        to form previously unknown functional receptors. The new subunit
        is expressed in the central and peripheral nervous systems and in
        PC12 cells.
SQ      Sequence 2460 BP; 519 A; 767 C; 587 G; 587 T;

Query Match
Best Local Similarity 6.5%; Score 44.6; DB 1; Length 2460;
Matches 71; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY      575 TGCATATTCGTATGACGCGTTTGACCCACATTCACACTACAGTGTGGTAAATCT 634
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      409 TACACACATGCGGATGACGACCATATGAGTGTCTCTACACACAGTATGTGGTCC 468
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
C-      635 TCTGGGATTCGCGAGTACCTCCCTCCAGGCAATATCAAGAGTCTCTACATCG 689
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
L-      469 AACGGCAGCATTCAGTGGTGGCCCTGCTCTATCTACAAAGAGTGGCTGCAAGATTG 523
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
T95785 standard; cDNA to mRNA; 2383 BP.
AC      T95785;
DT      22-MAY-1998 (first entry)
DE      Maize starch synthase type I cDNA
KW      Maize; starch synthase type I; starch; ds.
OS      Zea mays.
FH      Key Location/Qualifiers
FT      2..1951
FT      /*tag-a
FT      /product= starch_synthase_type_I
PN      DE19619918-A1.
PD      20-NOV-1997.
PF      17-MAY-1996; 019918.
PR      17-MAY-1996; DE-019918.
PA      (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
PI      Froberg C, Kossmann J;
DR      WPI; 98-000821/01.
DR      P-PSDB; W38218.
PT      DNA encoding maize starch synthase type I protein - for producing
PT      transgenic plants
PS      Claim 1; Pages 16-20; 23pp; German.

```

```

CC      The present sequence encodes maize starch synthase type I, useful
CC      in the production of starch. Starch can be used in various
CC      conventional starch applications, e.g. starch hydrolysate products,
CC      foods, papermaking, adhesives, textiles, building materials, soil
CC      briquettes, agrochemicals, pharmaceuticals, cosmetics, coal
CC      and synthetic polymers. The enzyme produces a starch stated to have
CC      different physicochemical properties, especially viscosity and
CC      gelling properties, from wild type starch.
SQ      Sequence 2383 BP; 562 A; 546 C; 633 G; 622 T;

Query Match
Best Local Similarity 6.2%; Score 42.8; DB 1; Length 2383;
Matches 83; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY      170 GCTCCACAGAGGTCACGGCGGGAGAGTGAGCCGAGAGCTGGCGGGGGCC 229
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      121 GCTCCACAGCGCTGCTCGCGCGGGGCTCCGAGAGTACAGGAGGCCCGCGCC 180
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      230 CGGCTGCGGCGCGCGCCATGACAGCGGCTCGGAGCTCTTTTCGCGGCCCTCC 289
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      181 GCGCCGATGCCACCGCGCTCTGCGCCCGCGCTCGTCCGCGGCTTCTCGCGCCG 240
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      290 CCGCGAGGTGAGGGGAGAGATGTCATGTC 319
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      241 GCGCGAGCCGACGGGTGAGCGCGCATTTGAC 270
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
X28103/c
ID      X28103 standard; DNA; 2106 BP.
AC      X28103;
DT      11-JUN-1999 (first entry)
DE      Frecac11 gene.
KW      Frecac11; transcription factor; adipose tissue expressed gene; obesity;
KW      lipid metabolism; adipocyte differentiation; obesity-related condition;
KW      non-insulin dependent diabetes mellitus; cardiovascular disease;
KW      catabolic condition; anorexia; bulimia; therapy; ss.
OS      Homo sapiens.
PN      W09854216-A1.
PD      03-DEC-1998.
PR      26-MAY-1998; SE0989.
PR      26-MAY-1997; SE-001963.
PA      (PMAA ) PHARMACIA & UPJOHN AB.
PI      Carlsson P, Enerbaeck S;
DR      WPI; 99-070211/06.
DR      P-PSDB; Y01097.
PT      New transcription factor designated FRECAC11 which regulates
PT      adipose tissue expressed genes involved in lipid metabolism and
PT      adipocyte differentiation, used to, e.g. inhibit gene expression
PS      Claim 1; Fig 1; 23pp; English.
        This sequence represents the Frecac11 gene of the invention. The encoded
        protein has transcription regulatory function directed against adipose
        tissue expressed genes. The protein is involved in lipid metabolism
        and/or adipocyte differentiation. FRECAC11 is selectively expressed in
        adipose tissue, and is used for transcriptional regulation of adipocyte
        obesity-related conditions to increase or decrease the activity of
        CC the cis elements of adipose expressed genes, e.g. FRECAC11 can be used as a drug that blocks
        CC antisense constructs of Frecac11 can be used to inhibit gene expression;
        CC The FRECAC11 protein can also be used to down regulate screening for
        CC substances that affect the activity of FRECAC11, such as inhibitors,
        CC antagonists, or agonists. Conditions which can be treated include
        CC obesity, non-insulin dependent diabetes mellitus, cardiovascular
        CC diseases, catabolic conditions, anorexia, bulimia.
SQ      Sequence 2106 BP; 419 A; 757 C; 602 G; 328 T;

Query Match
Best Local Similarity 6.0%; Score 41.2; DB 1; Length 2106;
Matches 109; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

```



CC constructs useful for gene activation by homologous recombination.  
 CC Novel genes are generated in which exogenous and endogenous exons  
 CC are operatively linked. Expression of IPO encoded by the endogenous  
 CC gene is activated, providing a means for the gene therapy of  
 CC thrombocytopenia.  
 SQ Sequence 4488 BP; 875 A; 1440 C; 1439 G; 722 T;

Query Match  
 Best Local Similarity 47.3%; Score 38.2; DB 1; Length 4488;  
 Matches 115; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 60 GACGAGAGCCGCTCTCTGCTGAGTGTATTAAGGAGCAGCCCGAGCCG 119  
 DB 2401 GCGGAGAGGCGCGGCGCGGCGGCGGCGGAGGCGCGGCGGCGGCGG 2460  
 QY 120 CCAATAGCTCCCGCCAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 179  
 DB 2461 GGGCAGTGTCCCGAGAGGCGCTGCTGCGGCGCGGCGGCGGCGGCGG 2520  
 QY 180 GGGTCAAGGCGGCGGAGAGTGTGAGCCCGAGAGCTGCGGCGGCGGCGG 239  
 DB 2521 CCGGAGAGAGGCGGCGGAGAGCGCGGCGGCGGCGGCGGCGGCGGCG 2580  
 QY 240 GCCGCGCCATGACAGCGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTG 299  
 DB 2581 CTACTGCCCCCGGCGCGGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTG 2640  
 QY 300 GAG 302  
 DB 2641 GGG 2643

## RESULT 15

Q61817  
 ID Q61817 standard; DNA; 1578 BP.

AC Q61817;  
 DT 11-NOV-1994 (first entry)  
 DE Open reading frame of pseudorabies virus.  
 KW Pseudorabies virus; antigen; polypeptide; vaccine; pig; swine;  
 KM disease; mutant; mutation; ds.  
 OS Pseudorabies virus.

FT Key Location/Qualifiers  
 FT cds 1..1578

FT /tag- a  
 FT /product- Antigenic or functional polypeptide of  
 FT pseudorabies virus.

FF W09408000-A.  
 PD 14-APR-1994.  
 PF 06-OCT-1993; E02738.  
 PR 06-OCT-1992; EP-203079.  
 PA (ALKU) AK20 NY.  
 PI Mettenleiter TC.  
 DR WPI; 94-135567/16.  
 DR P-PSDB; R51272.  
 PT New pseudo:rabies virus mutants - having a mutation in an open  
 PT reading frame used partic. in vaccines for immunising pigs.  
 PS Claim 1; Page 23-24; 33pp; English.  
 CC A pseudorabies virus with a mutation in this open reading frame can  
 CC be used in the production of vaccines for combatting infectious  
 CC diseases in pigs. The pseudorabies virus mutant produced fails to  
 CC produce the antigenic or functional polypeptide described in R51272  
 CC and expresses a polypeptide with altered antigenic or functional  
 CC properties or one that is non-functional. The mutation is either a  
 CC nucleic acid substitution, deletion, insertion or inversion or a  
 CC combination of these. The preferred substitutions are a combination  
 CC of substitutions selected from those where the nucleic acid is  
 CC modified so that the codon encoding His at position 37 is replaced  
 CC by one encoding Arg, the codon encoding Glu at position 355 is  
 CC changed to one encoding Asp, and the codon encoding Val at position  
 CC 375 is changed to a codon encoding Ala. A mutation in the control  
 CC elements of the ORF localised in an intergenic sequence may also  
 CC result in a failure of expression of the polypeptide.

SQ Sequence 1578 BP; 186 A; 568 C; 607 G; 217 T;

Query Match  
 Best Local Similarity 50.5%; Score 38; DB 1; Length 1578;  
 Matches 92; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 53 GCGGCTGAGCGGAGCCGCGCTCTGCTGAGTGTATTAAGGAGAGCCCG 112  
 DB 1019 GCGCGGCGGCGGCGGCGGCGGCGGCGGAGAGAGAGAGAGAGAGG 1078  
 QY 113 CAGCGCGCCATAGCTCCCGCCAGATCTCTGCTGCTGCTGCTGCTGCTG 172  
 DB 1079 GCGGCCAGCGGCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1138  
 QY 173 CCGACGAGGATCACGCGCGGAGAGAGTGTGAGCCCGAGAGCTGCGCGGCG 232  
 DB 1139 GCGAGCGAGACGCTTCTGCGGCGGAGGTGATGCGGATGCTGTGAACCG 1198  
 QY 233 CC 234  
 DB 1199 CC 1200

Search completed: September 18, 1999, 00:35:12  
 Job time: 19072 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:33 ; Search time 2825.05 Seconds  
(without alignments)  
447.565 Million cell updates/sec

Title: US-08-956-518A-103

Perfect score: 641 CAGGCCCCGACATAGCTGCC.....CAGAGTTCCTGCTACATCG 641

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	151.8	23.7	560	26	W52861	W52861 zc03e02.r1
2	130.8	20.4	409	25	W03952	W03952 za62c08.r1
3	106	16.5	170	25	N83152	N83152 K4329F Huma
4	92.2	14.4	510	54	HSW011948	A1047098 Homo sapi
5	76.4	11.9	487	39	AA861176	AA861176 ak36d09.s
6	68.6	10.7	335	32	AA333244	AA333244 EST37317
7	49.4	7.7	607	44	A1292581	A1292581 GH1518.5
8	46	7.2	501	49	A1655990	A1655990 t442c03.x
9	43.6	6.8	274	34	AA460018	AA460018 z449h09.s
10	43.6	6.8	436	39	AA844642	AA844642 a157a12.s
11	43.6	6.8	396	45	A1367541	A1367541 qv93c12.x
12	40.8	6.4	494	37	AA697326	AA697326 H02253.5
13	39.4	6.1	439	47	A1492967	A1492967 q47c02.x
14	39.2	6.1	354	26	W86937	W86937 zh60e03.r1
15	39	6.1	557	43	A1192683	A1192683 ge62d07.r
16	38.6	6.0	369	29	AA156428	AA156428 t215c02.s
17	37.8	5.9	790	48	A1595069	A1595069 vb77d03.y
18	37	5.7	473	33	AA407186	AA407186 EST02181
19	36.6	5.7	319	33	AA408177	AA408177 EST02759
20	36.6	5.7	433	40	AA923335	AA923335 O144e04.s
21	36.4	5.7	400	50	A1681298	A1681298 tx45f10.x
22	36.4	5.7	399	51	AU067986	AU067986 AU067986
23	36	5.6	435	42	A1085561	A1085561 cy68a04.x
24	36	5.6	344	45	A1361153	A1361153 qy05c04.x
25	36	5.6	373	47	A162417	A162417 ub62e06.x
26	36	5.6	373	47	A162417	A162417 ub62e06.x
27	35.6	5.5	415	24	H81299	H81299 yu74f03.s1
28	35.4	5.5	317	46	A1443300	A1443300 sa45b11.y
29	35.2	5.5	638	26	W27222	W27222 26b9 Human
30	35.2	5.5	654	45	A1346522	A1346522 gp51d09.x
31	35	5.5	232	20	D22064	D22064 RICC10276A
32	35	5.5	232	20	D22108	D22108 RICC10350A
33	35	5.5	216	20	D24865	D24865 RICC2674A.R
34	35	5.5	482	39	AA894793	AA894793 c155e11.s
35	35	5.5	482	40	AA60583	AA60583 we68d07.x
36	34.6	5.4	368	40	AA953883	AA953883 on76d09.s
37	34.4	5.4	328	33	AA407004	AA407004 EST02243
38	34.4	5.4	416	33	AA407437	AA407437 EST00734
39	34.4	5.4	417	43	A1198318	A1198318 q161d07.x
40	34.4	5.4	461	43	A1199677	A1199677 q160a10.x
41	34.4	5.4	351	45	A1364148	A1364148 q176f02.x
42	34.4	5.4	495	48	A1564797	A1564797 tn37d02.x
43	34.4	5.4	372	50	AU064505	AU064505 AU064505
44	34.2	5.3	328	24	H86162	H86162 ys94b03.r1
45	34.2	5.3	298	43	A1227041	A1227041 u131h12.y

## ALIGNMENTS

RESULT 1  
LOCUS W52861  
DEFINITION zc03e02.r1 Soares,Parathyroid\_tumor\_NHPR Homo sapiens cDNA clone  
IMAGE:321242 5 similar to gb:X70297 NEURONAL ACETYLCHOLINE  
RECEPTOR PROTEIN, ALPHA-7 CHAIN (HUMAN);, mRNA sequence.  
ACCESSION W52861  
NID g1350351

VERSION	W52861.1	GI:1350351
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
AUTHORS	Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 560) Hillier L., Clark N., Dubucq T., Elliston K., Hawkins M., Holman M., Hultman M., Kubacka T., Le M., Lennon G., Marra M., Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F., Trevaskis E., Waterston R., Williamson A., Wollmann P. and Wilson R.	
TITLE	The WashU-Merck EST Project	
JOURNAL	Unpublished (1995)	
COMMENT	On Jan 24, 1995 this sequence version replaced gi:634363.	
FEATURES	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 1735 Std Error: 0.00 Seq primer: mob.REGA+ET High quality sequence stop: 408. Location/Qualifiers 1..560 /organism="Homo sapiens" /db_xref="GDB:1258900" /db_xref="taxon:9606" /clone_id="IMAGE:321242" /clone_lib="Soares_parathyroid_tumor_NBHPA" /tissue_type="parathyroid tumor" /dev_stage="adult" /lab_host="DHIOB (ampicillin resistant)" /note="Organ: parathyroid gland; Vector: pRT73D (Pharmacia) with a modified polylinker; Site_1: Not I, Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15' TGTTACCACTGAGTAGGGGAGCGCCGCCACCATTATTTTTTTTTTTTTTTT T-3')", double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRT73D vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH.	
BASE COUNT	128 a 137 c 137 g 152 t	6 others
ORIGIN		
Query Match	23.7%	Score 151.8; DB 26; Length 560;
Best Local Similarity	96.4%	Pred. Mod. 5.1e-29;
Matches 186; Conservative	0; Mismatches 4;	Indels 3; Gaps 3;
OY 449	AATTGTGATGC AAAAATAATGTGATCTACACAGCATTTTCAGTTCCAAATGCTAATCCAGCAT	508
Db 1	ATTGTGATGC AAAAATAATGTGATCTACACAGCATTTTCAGTTCCAAATGCTAATCCAGCAT	60
OY 508	TTGTGGATAGCTGC AAATCGCATAATGCTATATGACCGCTTTGACGCCACTCCACT	568
Db 61	TTGTGGATAGCTGC AAACGATGATATGCTGATGACCGCTTTGAC-NCACATTCACACT	119
OY 569	AACGTTGGTGAATTCCTCTGGGCAATGCCAGTACCTCCGCCAGGCAATTCAGAAGT	628
Db 120	AACGTTGGTGAATTCCTCTCTGGGCAATTC-CAGTACTGCTCTCCA-GNATATTCAGAAGT	177
OY 629	TCCTGCTACTCATCG 641	

	DB	178	TCCGTGCTACATCG	190	 
	RESULT	2			
	LOCUS	W03952			
	DEFINITION	w03952 409 bp mRNA EST 19-Apr-1996 z662c08.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:297134.5 similar to gb:X70297 NEURONAL ACETYLCHOLINE RECEPTOR PROTEIN, ALPHA-7 CHAIN (HUMAN); mRNA sequence.			
	ACCESSION	W03952			
	NID	G1273817			
	VERSION	W03952.1			
	KEYWORDS	EST.			
	SOURCE	Homo sapiens human. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.			
	ORGANISM	Homo sapiens			
	REFERENCE	1 (bases 1 to 409) Hallier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin.L., Rohlfing,T., Soares,M., Tan,F., Trevaaks,E., Waterston,R., Williamson,A., Woldmann,P. and Wilson,R.			
	AUTHORS	The Masuh-Merck EST Project Unpublished (1995)			
	TITLE	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium ( <a href="http://infoimage.llnl.gov">infoimage.llnl.gov</a> ) for further information. Seq primer: mob.REGA+ER High quality sequence stop: 263. Location/Qualifiers			
	JOURNAL	1. 409 /organism="Homo sapiens" /db_xref="GDB:1242055" /db_xref="taxon:9606" /clone="IMAGE:297134" /clone_lib="Soares fetal liver spleen INFLS" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="DH10B (ampicillin resistant)" /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site.1: Pac I; Site.2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' ACTCGAGAAATTAATTAAAGACTCTTTTTTTTTTTTTT 3'] , double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."			
	FEATURES				
	SOURCE				
	BASE COUNT	95 a 80 c 75 g 145 t 14 others			
	ORIGIN				
	Query Match	20.4%; Score 130.8; DB 25; Length 409;			
	Best Local Similarity	92.4%; Pred. No. 1.le-23;			
	Matches 146; Conservative	0; Mismatches 11; Indels 1; Gaps 1;			
OY	482	TTTCGATTCGAATTCCTAATCCA-GCAATTTGGTAGTGCTGCAGAACGTGGCATTTGCTGA	540		
Dd	153	TTCTGATTCGAATTCCTAATCCACCATTNTNGATAGCTGCCAACTGGCATTTGCTGA	212		
OY	541	TGAGGCGTTTGAGCGCACATTCACACAAGAATACTGTTGGTGAATTCCTTGGGCATTTGCCA	600		
Dd	213	TGAGGCGTTTGAGCNCACACATTCACACACTRACAGTGTGGTGGAATTCCTTGGGCATTTGCCA	272		
OY	601	GTACCTGCCTCCAGGCATATTCAGAAGTTCCGTGCTACA	638		





ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 487)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 9, 1998 this sequence version replaced gi:937098.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/FLMN at: [www.bio.lnl.gov/bbrp/image/image.html](http://www.bio.lnl.gov/bbrp/image/image.html)  
Insert Length: 758 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 458.  
Location/Qualifiers  
1. 487  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1408049"  
/clone\_lib="Soares\_testis\_NHT"  
/sex="male"  
/lab\_host="DH10B"  
/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5', TGTTCCCATCTGAAAGTGGAGCGGCCCAATTTTCTTTTCTTTT 3']  
Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library went through one round of normalization to Cots, and was constructed by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 131 a 95 c 102 g 159 t  
ORIGIN  
Query Match 11.9%; Score 76.4; DB 39; Length 487;  
Best Local Similarity 98.7%; Pred. No. 8.7e-10;  
Matches 77; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 457 GCAGAAATATGTCATTCACAGCATTTTCAGTTCATTCATTCACAGCATTTGAT 516  
DB 487 GCAGAAATATGTCATTCACAGCATTTTCAGTTCATTCATTCACAGCATTTGAT 428  
QY 517 AGCTGCAACTGCGATAT 534  
DB 427 AGCTGCAACTGCGATAT 410

REFERENCE 1 (bases 1 to 335)  
AUTHORS Adams,M.D., Keilavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gockyne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodex,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.N., Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y., Badnarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-P., Fertile,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.  
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl.), 3-174 (1995)  
96026280  
JOURNAL MEDLINE  
COMMENT On May 18, 1995 this sequence version replaced gi:811121.  
Other ESTs: THC185696  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: [arke@veti.gr.org](mailto:arke@veti.gr.org)  
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1. 335  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):134920"  
/db\_xref="taxon:9606"  
/clone\_lib="Embryo, 8 week I"  
/dev\_stage="Embryo, 8 wks"  
/note="Organ: Embryo, 8 weeks; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"  
BASE COUNT 84 a 94 c 88 g 69 t  
ORIGIN  
Query Match 10.7%; Score 68.6; DB 32; Length 335;  
Best Local Similarity 84.6%; Pred. No. 7.5e-08;  
Matches 77; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
QY 175 TCCGCGGAGGTAGGGGAGATGTCATGTCAGGGTCAAGGCCAAGCAAGTTACT 234  
DB 1 TCCGCGGAGATAGGGGAGATGTCGTCAGGGCTCAAGGCCAGCTGAAGTTCT 60  
QY 235 GGCCTCATCTTCCAGGAGACCAAGAGCCA 265  
DB 61 GCGTCATCTTCAGCAAGACCAAGAGCCA 91

RESULT 7  
LOCUS A1292581 607 bp mRNA EST 01-DEC-1998  
DEFINITION GH15518, Sprime GH Drosophila melanogaster head pot2 Drosophila melanogaster cDNA clone GH15518 Sprime, mRNA sequence.  
ACCESSION A1292581  
NID 93941988  
VERSION A1292581.1 GI:3941988  
KEYWORDS EST.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster

REFERENCE 1 (bases 1 to 607)  
 AUTHORS Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,  
 Brokstein, P., Lewis, S. and Rubin, G.M.  
 TITLE BDGP/HMM Drosophila EST Project  
 JOURNAL Unpublished (1997)  
 COMMENT On Jun 18, 1996 this sequence version replaced gi:1366631.

FEATURES  
 source  
 Contact: Harvey, D.  
 G. M. Rubin-Molecular and Cell Biology  
 University of California Berkeley  
 539 USA, Berkeley, CA 94720-3200, USA  
 Fax: 510 643 9947  
 Email: <http://www.fruitfly.org/EST/estefruitfly.berkeley.edu>  
 hit genomic sequence AC005890  
 Plate: 155 row: B column: 6  
 High quality sequence stop: 465.  
 Location/Qualifiers  
 1..607  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone="GH15518"  
 /clone\_11b="GH Drosophila melanogaster head POT2"  
 /sex="male and female"  
 /dev\_stage="adult"  
 /lab\_host="DH5 - alpha"  
 /note="organ: head; Vector: POT2; Site:1; EcorI: Site:2;  
 XhoI: Sized fractionated cDNAs were directly ligated into  
 POT2. Plasmid cDNA library."

BASE COUNT 157 a 151 c 162 g 137 t  
 ORIGIN

Query Match 7.7%; Score 49.4; DB 44; Length 607;  
 Best Local Similarity 64.3%; Pred. No. 0.0073;  
 Matches 74; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 527 TGGCATATTGCTGATGAGCGTTGACGCACATTCACACTAACGTTGTAATTC 586  
 Db 398 TACACAGCGCGGATGAGGATTCGATGCGACATACACACCGCTTGCGCAACAT 457  
 QY 587 TGTGGCATTCGCCAGTACCTGCTCCAGCATATTCAGAGATTCCGTACATCG 641  
 Db 458 GCGCGAGTGTCTGTACGTGCGCCCTGCTATCTCAAGACGACATCAAGATGG 512

LOCUS 8  
 .653990/c  
 DEFINITION A1655990 501 bp mRNA EST 04-MAY-1999  
 t142c03.x1 NCI\_CGAP\_GC6 Homo sapiens cDNA clone IMAGE:2243428 3',  
 mRNA sequence.  
 ACCESSION A1655990  
 NID G4739968  
 VERSION A1655990.1 GI:4739969  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 501)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3188057.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: [Robert.Strausberg@nih.gov](mailto:Robert.Strausberg@nih.gov)  
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
 R. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bernaldo, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
[www.bio.lnl.gov/btrp/image/image.html](http://www.bio.lnl.gov/btrp/image/image.html)

Seq primer: -400p from glbco  
 High quality sequence stop: 458.  
 Location/Qualifiers  
 1..501  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2243428"  
 /clone\_11b="NCI\_CGAP\_GC6"  
 /tissue\_type="pooled germ cell tumors"  
 /lab\_host="DH10B"  
 /note="vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; Plasmid DNA from the normalized library  
 NCI-CGAP\_GC4 was prepared, and ss circles were made in  
 vitro. Following HAP purification, this DNA was used as  
 tracer in a subtractive hybridization reaction. The driver  
 was PCR-amplified cDNAs from a pool of 5,000 clones made  
 from the same library (cloneids 1257096-1258631,  
 1469064-1470983, and 1475592-1476743). Subtraction by  
 Bento Soares and M. Fatima Bernaldo. "

BASE COUNT 135 a 107 c 115 g 144 t  
 ORIGIN

Query Match 7.2%; Score 46; DB 49; Length 501;  
 Best Local Similarity 83.9%; Pred. No. 0.051;  
 Matches 52; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 427 ATGAAACACCCACCGTGAATTTGATGCAAAATTTGCATCTACACATTTCA 486  
 Db 499 ATATGAAACACCCACCGTGAATTTGATGCAAAATTTGCATCTACACATTTCA 440  
 QY 487 GT 488  
 Db 439 GT 438

RESULT 9  
 LOCUS AA460018/c  
 DEFINITION AA460018 274 bp mRNA EST 09-JUN-1997  
 zx49h03.s1 Soares\_testis\_NHT Homo sapiens cDNA clone IMAGE:795617  
 3', mRNA sequence.  
 ACCESSION AA460018  
 NID G2184902  
 VERSION AA460018.1 GI:2184902  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 274)  
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,  
 Kueba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J.,  
 Moore, B., Scheinberg, R., Steptoe, M., Tan, F., Theising, B.,  
 White, Y., Wyllie, T., Waterston, R. and Wilson, R.  
 TITLE WASHU-Merck EST Project 1997  
 JOURNAL Unpublished (1997)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1393691.

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: [estevan@wustl.edu](mailto:estevan@wustl.edu)  
 This clone is available royalty-free through LNL; contact the



Site 2: Not: Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.85 kb. Life Technologies catalog #:  
11539-012"

BASE COUNT 98 a 86 c 97 g 115 t  
ORIGIN

Query Match 6.8%; Score 43.6; DB 45; Length 396;  
Best Local Similarity 92.0%; Pred. No. 0.19;  
Matches 46; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 485 CAGTCCATTGCTATCCAGCATTTGTGGATGCGCAACGCGATAT 534  
1 |||||  
Db 280 CTGTCATATGCTATCCAGCATTTGCGATAGCTCCAACTGGGATAT 231

RESULT 12  
AA697326 494 bp mRNA EST 28-NOV-1998  
LOCUS HL02253.5prime HL Drosophila melanogaster head Bluescript  
DEFINITION Drosophila melanogaster cDNA clone HL02253 5prime similar to  
M20316: D.melanogaster acetylcholine receptor-related protein mRNA,  
complete cds, mRNA sequence.

ACCESSION AA697326  
NID 92700255  
VERSION AA697326.1 GI:2700255  
KEYWORDS EST.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster

REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 494)  
AUTHORS Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,  
Brostein, P., Lewis, S. and Rubin, G.M.  
BDGP/HMI Drosophila EST Project  
TITLE Unpublished (1997)  
JOURNAL On Nov 20, 1997 this sequence version replaced gi:1172316.  
COMMENT

FEATURES  
source  
1. .494  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="HL02253"  
/clone\_1lb="HL Drosophila melanogaster head Bluescript"  
/sex="male and female"  
/dev\_stage="adult"  
/lab\_host="SOLR"  
/note="Organ: head-brain & sensory organ. Vector:  
Bluescript SK; Site 1: EcoRI; Site 2: XhoI; Constructed  
using Stratagene ZAP-cDNA Synthesis kit. Oligo dt-primed  
and directionally cloned at EcoRI and XhoI in Bluescript  
SK(+/-)."  
Location/Qualifiers  
High quality sequence stop: 416.

BASE COUNT 124 a 114 c 127 g 129 t  
ORIGIN

Query Match 6.4%; Score 40.8; DB 37; Length 494;  
Best Local Similarity 59.5%; Pred. No. 1.1;  
Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 526 CTGCGATATGCTGATGACGCTTTGACGCGACATTCACACTACGCTGTGGTGAATTC 585  
1 |||||  
Db 244 CTGATATATGCGGATGCGACACTACGAGGTGGCGCTACAGTCCAGCTGCTATTAATCC 303

QY 586 TTCTGGCATTCGCCAGTACCTGCTCCAGGCATATTCAGAGCTTCTGTACATCG 641  
1 |||||  
Db 304 CACGGGAGAGGCTCTGTGGTCTCTCCGGCCATTATACAGAGCTCTGCACATCG 359

RESULT 13  
A1492967 439 bp mRNA EST 18-MAR-1999  
LOCUS q247c02.x1 NCI-CGAP\_K1d11 Homo sapiens cDNA clone IMAGE:2030018 3'  
DEFINITION similar to SW:CDNC\_HUMAN P49918 CYCLIN-DEPENDENT KINASE INHIBITOR  
1C; contains TAR1.t2 TAR1 repetitive element; mRNA sequence.

ACCESSION A1492967  
NID 94393970  
VERSION A1492967.1 GI:4393970  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 439)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On May 18, 1998 this sequence version replaced gi:3136536.

FEATURES  
source  
1. .439  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="21q22.1-3"  
/clone="IMAGE:2030018"  
/clone\_1lb="NCI-CGAP\_K1d11"  
/lab\_host="DH10B"  
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site 1: Not I; Site 2: Eco RI;  
Plasmid DNA from the normalized library NCI-CGAP Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(clonids 132376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."  
www.bio.livl.gov/dbtrp/image/image.html

Insert Length: 908 Std Error: 0.00  
Seq primer: -40UP from Glibco  
High quality sequence stop: 65.  
Location/Qualifiers  
1. .439

BASE COUNT 60 a 145 c 187 g 47 t  
ORIGIN

Query Match 6.1%; Score 39.4; DB 47; Length 439;  
Best Local Similarity 53.3%; Pred. No. 2.3;  
Matches 105; Conservative 0; Mismatches 91; Indels 1; Gaps 1;

QY 26 GTCTCGGTGCCCCCTTCATTTCCAGCCGCTCCACGAGGCTCAGCGGCGGGA 85  
1 |||||  
Db 46 GGCAGCGGCGCCCGCTGTTGGCGCCCTGCTGCTTGAAGCCCGCGGTCCGGGG 105  
QY 86 GAGGTGAGCCCGAGAGCTCGCGGGGGCCCGCTGTGTGGCGCGCATGACAGCG 145  
1 |||||

Db	106	CCGAGGCGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGT-GCTGAGGACCGGGGCGCG	164
Qy	146	GCTCGGAGCTAGGCTCTCTTTTCGCGGCGCCCTCCCGCGGAGGAGGAGGAAAGATTCATG	205
Db	165	ACTGAGACCGGGGCGGAGCGGAGCGGAGCGCGGGGCGCGGGCCAGACCAAG	224
Qy	206	TCAGGGTTCAAGGCCAA	222
Db	225	ACCGGACCGGAGCGGA	241

RESULT 14					
W86937					
LOCUS	W86937	354 bp	mRNA	EST	02-FEB-1997
DEFINITION	zh60e03.r1 Soares_fetal_liver.spleen.mf1LS.S1 Homo sapiens cDNA				
ACCESSION	clone IMAGE:416476 5', mRNA sequence.				
	w86937				

SION W86937.1 GI:1400676  
WORDS EST.

SOURCE	ORGANISM
human.	Homo sapiens

REFERENCE	1 (bases 1 to 354)
AUTHORS	Hallier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,

**AUTHORS**  
 Hallier, L., Lennon, G., Becker, M., Bernaldo, M.F., Chiappelli, B.,  
 Chissoe, S., Dietrich, N., Dubouquet, T., Favello, A., Gish, M.,  
 Hawkins, M., Hultman, M., Kuehba, T., Lacey, M., Le, N.,  
 Mards, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
 Trevasakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
 and Marra, M.

**TITLE** Generation and analysis of 280,000 human expressed sequence tags  
**JOURNAL** Genome Res. 6 (9), 807-828 (1996)  
**INDEXING** 02041460

MEDLINE	9/0444/8
COMMENT	On Apr 14, 1993 this sequence version replaced gi:315330

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [estewatson.wustl.edu](mailto:estewatson.wustl.edu)  
This clone is available royalty-free through LNL ; contact the  
IMAG Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
Insert Length: 584    Std Error: 0.00  
Seq primer: mob.REGA+ET  
High quality sequence stop: 217.

FEATURES	Location/Qualifiers
source	1. .354

```

/organism="Homo sapiens"
/db_xref="GDB:1324946"
/db_xref="taxon:9606"
/clone_image="IMAGE:416475"
/clone_lib="Soares_fetal_liver_spleen_INFLS_SI"
/sex="male"
/deriv_stage="20 week post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the Original Soares fetal
liver spleen INFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer 15'
AACGCGAGAGATTATATTAAGATCTCTTTTTTTTTTTTTTTT 3'
double-stranded cDNA was ligated to Eco RI adapters
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

```

Query Match	6.1%	Score 39.2	DB 26	Length 354
Best Local Similarity	75.3%	Pred. No. 2.5		
Matches 61	Conservative 0	Mismatches 19	Indels 1	Gaps 1

Qy	545	CGCTTGGACGCACATTCAC	565
Db	145	CGCTGGGTCTCACCCTTCAC	165
Qy	485	CAAGTCAATGATCTCAATCCAGCATTGTGTGACCTCAAACTGCATATGTGCATGAG	544
Db	86	CTGTCTCAATGATCTCAATCCAGTATTTCGGGATA-NTCCAAACTGGGATATGTCTCAGG	144

RESULT	15				
A1192663					
LOCUS	A1192663	557 bp	mRNA	EST	29-OCT-1998
DEFINITION	qec6d07.x1 Soares_fetal_lung_NbHL10W				
					Homo sapiens CDNA clone

IMAGE:1/743565 3' similar to SW:CDNC\_HUMAN P49918 CYCLIN-DEPENDENT KINASE INHIBITOR 1C ; contains MER22.b3 TARI repetitive element ; mRNA sequence.

ACCESSION	A1192683
NID	93743892
VERSION	A1192683.1
	GI:3743892

KEYWORDS	EST,
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE  
1 (bases 1 to 557)  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

**AUTHORS** NCI-CGAP <http://www.nci.nlm.nih.gov/ncicgap>.  
**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
**JOURNAL** Unpublished (1997)

JOURNAL Unpublished (1997)  
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797633

Contact: Robert Strausberg, Ph.D.  
 Tel.: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LINT ; contact the  
 IMAG Consortium (info@image.lint.gov) for further information.  
 Insert Length: 862 Std Error: 0.00  
 Seq primer: -40UP from Glibco  
 High quality sequence scop: 118.  
 Location/Qualifiers  
 1..557  
 FEATURES  
 source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1743565"
/clone_lib="Soares-Fetal_Lung_NbH19W"
/dev_stage="19 weeks"

```

```

BASE COUNT      62 a      193 c      242 g      60 t
ORIGIN
/ab.host="DH10B (ampicillin resistant)"
/notes="Organ: lung; Vector: pRT73 (pharmacina) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - Oligo(dT) primer
15'-GTGACCAATCTGAAGTGGAGCGCCGCAATTTTTTTTTTTTTTTT-3' (p
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacina), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pRT73 vector
(Pharmacina). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NBHH19W."

```

Query Match	6.1%	Score 39;	DB 43;	Length 557;
Best Local Similarity	53.3%	Pred. No. 3.2;		
Matches 104;	Conservative	0;	Mismatches 90;	Indels 1;
			Gaps	1.

26 GTCTCTGGGTGCCCCCTTGGCCATTTCACAGCCGCGCTCCACAGAGGGTACAGGGCGGGGGA 85

Db 68 GGGCCGGCTGCCCTGTGTCGGGCCCTGCTCGGCGCTCTGTGAGGCGCCGGCTCCGGG 127  
QY 86 GAGGTGGAGCCCGGAGAGCTCGGCGGGGGCCCGCTGTGGCCCGGCATGACAGCG 145  
Db 128 CCGGGCCCGGGCGGGGCGCGGGCCCGGGCGCG -GGCTGGGCGCGGGCGCGCG 186  
QY 146 GGTGGGAGCTGCTCTTTCCGCGGCCCTCCGCGCGGAGTGAGGGGAAGATGTCCATG 205  
Db 187 ACTGGAGCCGGGCGGAGCCGAGCCGGGCGGGCGCGGCGCAAGAGCCGCG 246  
QY 206 TCAGGTTCAAGCC 220  
Db 247 ACCGCGACCGAGCC 261

Search completed: September 17, 1999, 21:28:34  
Job time: 14298 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:35:12; Search time 425.19 Seconds

(without alignments)  
377.180 Million cell updates/sec

Title: US-08-956-518A-103

Perfect score: 641  
Sequence: 1 CAGGCCGCCACATAGCTCC.....CAGAGTTCCTGCTACATCG 641

Scoring table: IDENTITY\_NUC

arched: 311585 segs, 125096042 residues

\_cabase: N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	16.8	1876	1 T48239	Neuronal nicotinic
2	108	16.8	1876	1 V12197	Human neuronal nic
3	108	16.8	1590	1 V44687	V274T variant huma
4	74.4	11.6	2769	1 T59196	Neuronal alpha-bun
5	61.6	9.6	2101	1 T59197	Neuronal alpha-bun
6	44.6	7.0	2460	1 Q06086	Plasmid pAPC13 enc
7	43	6.7	11955	1 X53491	Human adenovirus A1
8	42.8	6.7	2383	1 T95785	Maize starch synth
9	41.2	6.4	2106	1 X28103	Freacell gene. New
10	40.2	6.3	2992	1 T67285	Soluble starch syn
11	40.2	6.3	2990	1 V66832	Zea mays soluble s
12	36.6	5.7	1915	1 T48241	Neuronal nicotinic
13	36.4	5.7	5561	1 V68408	Human BAZ1-beta CD
14	35.4	5.5	1942	1 Q84888	Human BAZ1-beta CD
15	35	5.5	1915	1 V12198	DNA encoding cytop
16	35	5.5	2538	1 X06987	Human neuronal nic
17	35	5.5	2538	1 V12196	Mouse neuregulin r
18	34.6	5.4	2363	1 V12196	Human neuronal nic
19	34.4	5.4	201	1 N70195	Streptomyces prote
20	34.4	5.4	1536	1 Q44491	Therinus aquaticus
21	34.4	5.4	4488	1 T35520	Human thrombopoiet
22	34.4	5.4	3496	1 T48237	Neuronal nicotinic
23	34.4	5.4	1925	1 T48234	Neuronal nicotinic
24	34.4	5.4	2082	1 T59527	Alpha4 subunit of
25	34.4	5.4	2082	1 T59528	Alpha4 subunit of
26	34.4	5.4	1120	1 V60339	CDNA sequence of f
27	34	5.3	3223	1 Q62917	hmrnp u protein co
28	34	5.3	774	1 T44499	Human cyclin-depen
29	33.8	5.3	1018	1 T02326	Human cardiac hype
30	33.8	5.3	1018	1 T87014	Human cardiolipol
31	33.8	5.3	43280	1 T80413	Tyrosine synthase
32	33.6	5.2	2721	1 Q10212	BamHI J-I fragment
33	33.6	5.2	2721	1 Q10543	BamHI J-I fragment
34	33.6	5.2	3000	1 Q14319	Pectactin antigen
35	33.6	5.2	3000	1 Q34566	Pectactin antigen
36	33.6	5.2	8438	1 Q73500	prn gene. DNA enco
37	33.6	5.2	1597	1 T12908	DNA encoding Pseud
38	33.6	5.2	4739	1 T78203	Human insulin rece
39	33.6	5.2	8169	1 V26609	Physiologically ac
40	33.4	5.2	201	1 N70194	Actinomadura hibis
41	33.4	5.2	2064	1 O52638	Signal portion of
42	33.4	5.2	2821	1 T83389	Streptomyces fradi
43	33.4	5.2	2821	1 T83389	Human cystatin B g
					Human mutant EPM1

## ALIGNMENTS

C 44 33.4 5.2 2821 1 T89395  
45 33.4 5.2 1320 1 V44138  
Mutant EPM1 gene a  
Chimeric rhpGf-B

RESULT 1

T48239 T48239, standard; DNA; 1876 BP.

AC T48239;

DE Neuronal nicotinic acetylcholine receptor alpha-7 subunit DNA.

KW Neuronal nicotinic acetylcholine receptor; nAChR; neurotransmitter;

OS Homo sapiens.

FH Key Location/Qualifiers

FT cds 73..1581

PI WO9641876-A1.

PD 27-DEC-1996.

PF 07-JUN-1996; U09775.

PR 07-JUN-1995; US-484722.

PA (SIBI-) SIBIA NEUROSCIENCES INC.

PI Eliott KJ, Harpold KM;

DR WPI; 97-065463/06.

DR P-PSDB; W09025.

PT Nucleic acids encoding neuronal nicotinic acetylcholine receptor sub-units -

PS used in screening to determine the effect of drugs on the receptor

PS Disclosure; Page 71-73; 108pp; English.

CC A DNA sequence (T48239) codes for the alpha-7 subunit (W09025) of

CC the human neuronal nicotinic acetylcholine receptor (nAChR). Host

CC cells, esp. mammalian cells or amphibian oocytes, carrying alpha-7

CC nucleic acids, opt. in combination with other alpha and/or beta

CC subunit nucleic acids (see also T48232-38, T48240-41), express

CC recombinant nAChR subunits useful for identifying cpds. that

CC modulate the activity of human nAChRs.

SO Sequence 1876 BP; 369 A; 553 C; 531 G; 423 T;

Query Match 16.8%; Score 108; DB 1; Length 1876;

Best Local Similarity 95.7%; Pred. No. 9.1e-23;

Matches 11; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 526 CCGCATATTTGCTGATGAGCGTTGAGCCACATTCACACTACGTTGGTGAATC 585

DB 414 CTATACAGCTGTGATGAGCGTTGAGCCACATTCACACTACGTTGGTGAATC 473

QY 586 TTCTGGGATTCGCCAGTACCTGCTCCAGGATATTCAGAGTTCTCTCATATCG 641

DB 474 TTCTGGGATTCGCCAGTACCTGCTCCAGGATATTCAGAGTTCTCTCATATCG 529

RESULT 2

V12197 V12197, standard; cDNA; 1876 BP.

AC V12197;

DE Human neuronal nicotinic acetylcholine receptor alpha-7 subunit cDNA.

KW Human; neuronal nicotinic acetylcholine receptor; alpha-7 subunit;

OS Homo sapiens.

FH Key Location/Qualifiers

FT cds 73..1581

PI WO9420617-A2.

PD 15-SEP-1994.

PF 08-MAR-1994; U02447.

PR 08-MAR-1993; US-028031.

PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.

(SIBI-) SIBIA NEUROSCIENCES INC.



PI Elliott KJ, Ellis SB, Harpold NM;  
 DR WPI: 94-303024/37.  
 DR P-PSDB: W44153.  
 PT Human neuronal nicotinic acetylcholine receptor subunits and DNA -  
 CC also transformed cells useful for screening cpds, which modulate  
 PS activity of the receptor  
 CC Claim 8; Page 78-79; 99pp; English.  
 CC The present sequence encodes a human neuronal nicotinic acetylcholine  
 CC receptor (NACHr) subunit. The cells expressing the alpha and/or beta  
 CC NACHr subunits may be used in a method of screening compounds to  
 CC identify any which modulate the activity of human neuronal NACHr.  
 CC Subunit specific antibodies may be used to monitor the distribution  
 CC and expression density of various subunits in normal vs diseased brain  
 CC tissues. Testing of single receptor subunits or specific receptor  
 CC subunit combinations with a variety of potential agonists or antagonists  
 CC provides information with respect to the function and activity of the  
 CC individual subunits and should lead to the identification and design of  
 CC compounds that are capable of very specific interaction with one or  
 CC more receptor subtypes. The resulting drugs should exhibit fewer  
 CC unwanted side effects than drugs identified e.g. screening with cells  
 CC that express a variety of subtypes  
 CC Sequence 1876 BP; 369 A; 533 C; 530 G; 423 T;  
 Query Match 16.8%; Score 108; DB 1; Length 1876;  
 Best Local Similarity 95.7%; Pred. No. 9.1e-23;  
 Matches 111; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 526 CTGCGATATGTGATGAGCGCTTTGACGCCACATTCACACTACAGTGTGGTGAATTC 585  
 DB 414 CTATACAGTGTGATGAGCGCTTTGACGCCACATTCACACTACAGTGTGGTGAATTC 473  
 QY 586 TTCTGGGCAATTCGACGACTGCTGCTCCAGGCAATTCACAGAGTCTCTGATACG 641  
 DB 474 TTCTGGGCAATTCGACGACTGCTGCTCCAGGCAATTCACAGAGTCTCTGATACG 529  
 RESULT 3  
 V44687  
 ID V44687 standard; cDNA; 1590 BP.  
 AC V44687;  
 DT 09-OCT-1998 (first entry)  
 DE V274T variant human alpha7 NACHr coding sequence.  
 KW Alpha7 NACHr; alpha7 nicotinic acetylcholine receptor subunit; cancer;  
 KW neurodegeneration; enzyme dysfunction; affective disorder; therapy;  
 KW immune dysfunction; diabetic neuropathy; Alzheimer's disease;  
 KW schizophrenia; ss.  
 OS Homo sapiens.  
 CS Key Location/Qualifiers  
 CDS 9..1517  
 /tag= a  
 PI W09828331-A2.  
 PN 02-JUL-1998.  
 PR 22-DEC-1997; U23405.  
 PR 20-DEC-1996; US-771737.  
 PA (ABBO) ABBOTT LAB.  
 PI Briggs CA, Gopalakrishnan M, McKenna DG, Monteggia LM,  
 PI Roch J, Sullivan JP, Touma E;  
 DR WPI: 98-377593/32.  
 P-PSDB: W69216.  
 PT Nucleic acid encoding variant of human alpha7 nicotinic  
 PT acetylcholine receptor sub-unit - used to identify modulators of  
 CC the receptor, potentially useful for treating neuro-degeneration,  
 CC cancer, affective disorders etc.  
 PS Claim 14; Fig 2; 44pp; English.  
 CC This sequence encodes the V247T variant of human alpha7 nicotinic  
 CC acetylcholine receptor (NACHr) subunit of the invention. Cells containing  
 CC the DNA are used to express the protein and to identify modulators of  
 CC alpha7 NACHr activity or cytoprotective agents, e.g. antisense  
 CC compounds or antagonists that are potentially useful for treating  
 CC neurodegeneration, enzyme dysfunction, affective disorders and immune  
 CC dysfunction, such as cancer, post-herpetic neuralgia, diabetic  
 CC neuropathy, osteoarthritis, Alzheimer's or Parkinson's diseases, kuru,

CC psychosis and schizophrenia. Probes based on the DNA are used to detect  
 CC the DNA in usual hybridisation or amplification tests, while monoclonal  
 CC antibodies are used to detect the protein for diagnosis (in vitro or by  
 CC in situ immuno-fluorescent assay). Compared with wild-type alpha7 NACHr,  
 CC the protein has about 100-fold greater sensitivity to cholinergic  
 CC receptor agonists (nicotine or acetylcholine) and response to these  
 CC agonists decays more slowly, but the wild-type inward rectification is  
 CC retained.  
 CC Sequence 1590 BP; 322 A; 471 C; 449 G; 348 T;  
 Query Match 16.8%; Score 108; DB 1; Length 1590;  
 Best Local Similarity 95.7%; Pred. No. 8.4e-23;  
 Matches 111; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 526 CTGCGATATGTGATGAGCGCTTTGACGCCACATTCACACTACAGTGTGGTGAATTC 585  
 DB 350 CTATACAGTGTGATGAGCGCTTTGACGCCACATTCACACTACAGTGTGGTGAATTC 409  
 QY 586 TTCTGGGCAATTCGACGACTGCTGCTCCAGGCAATTCACAGAGTCTCTGATACG 641  
 DB 410 TTCTGGGCAATTCGACGACTGCTGCTCCAGGCAATTCACAGAGTCTCTGATACG 465  
 RESULT 4  
 T59196  
 ID T59196 standard; cDNA; 2769 BP.  
 AC T59196;  
 DT 17-JUN-1997 (first entry)  
 DE Neuronal alpha-bungarotoxin binding protein alpha1 subunit cDNA.  
 KW Neuronal alpha-bungarotoxin binding protein alpha 1; cholinergic;  
 KW ligand binding; ion channel; ss.  
 OS Gallus sp.  
 CS Key Location/Qualifiers  
 CDS 71..1513  
 /tag= a  
 FT signal\_peptide 71..136  
 FT mat\_peptide 137..1510  
 FT /tag= b  
 FT /tag= c  
 PN US559709-A.  
 PD 04-FEB-1997.  
 PR 28-SEP-1989; 413947.  
 PR 28-SEP-1989; US-413947.  
 PA (SALK) SALK INST BIOLOGICAL STUDIES.  
 PI Lindstrom JM, Schoepfer RD;  
 DR WPI: 97-118297/11.  
 P-PSDB: W12368.  
 DT New isolated neuronal alpha-bungarotoxin-binding protein DNA - used  
 PT to screen cholinergic agents and other drugs which may affect ligand  
 PT binding, ion channel or other activities of the protein.  
 PS Claim 1; Fig 2A-B; 18pp; English.  
 CC 2 cDNA clones (T59196 and T59197) respectively code for the alpha1  
 CC alpha-bungarotoxin binding protein (ABBP). They were isolated from  
 CC an 18-day embryo chick brain cDNA library using a probe (see also  
 CC T59198) based on the N-terminal amino acid sequence of chicken  
 CC brain ABBP. The probe isolated partial clone pCh29-1, which  
 CC encoded the N-terminal portion of alpha1. A subclone, pCh29-3  
 CC (ATCC 40641), was used to rescreen the library, yielding clone  
 CC pCh31-1 (ATCC 40640), which encoded the entire alpha2 sequence.  
 CC A probe based on the C-terminal region of pCh31-1 was used obtain  
 CC clone pCh34-1 (ATCC 40639), encoding the C-terminal portion of  
 CC alpha1. The cDNA clones can be used as probes to identify further  
 CC ABBP subunits, and in the recombinant prodn. of ABBP.  
 CC Sequence 2769 BP; 790 A; 604 C; 571 G; 804 T;  
 Query Match 11.6%; Score 74.4; DB 1; Length 2769;  
 Best Local Similarity 77.6%; Pred. No. 1.4e-12;  
 Matches 90; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
 QY 526 CTGCGATATGTGATGAGCGCTTTGACGCCACATTCACACTACAGTGTGGTGAATTC 585

```

DB 412 CTATACACAGTCTGATGAAAGATTATGCTACATTCACACATTAAGTTTACTCAATTC 471
OY 586 TTCGGGACATTCGACATGACCTGCTCCAGCAGCATATTCAAGAGTCTGCTACATCG 641
DB 472 TTCGGGACATTCGACATGACCTGCTCCAGCAGCATATTCAAGAGTCTGCTACATAG 527

RESULT 5
T59197
ID T59197 standard; cDNA; 2101 BP.
AC T59197;
DT 17-JUN-1997 (first entry)
DE Neuronal alpha-bungarotoxin binding protein alpha2 subunit cDNA.
KM Neuronal alpha-bungarotoxin binding protein alpha 2; cholinergic;
OS Gallus sp.
Key cds Location/Qualifiers
  signal_peptide 56..145
  mat_peptide 146..1498
  *tag= a
  *tag= b
  *tag= c
  *tag= C

US559709-A.
PD 04-FEB-1997.
PF 28-SEP-1989; 413947.
PR 28-SEP-1989; US-413947.
PA (SALK ) SALK INST BIOLOGICAL STUDIES.
PI Lindstrom JM. Schoepfer RD.
DR WPI: 97-118297/11.
P-PSDB: W12369.
PT New isolated neuronal alpha-bungarotoxin-binding protein DNA - used
PT to screen cholinergic agents and other drugs which may affect ligand
PT binding, ion channel or other activities of the protein.
PS Claim 1; Fig 3A-B; 18pp; English.
CC 2 cDNA clones (T59196 and T59197), respectively code for the alpha1
CC subunit (W12368) and alpha2 subunit (W12369) of chick neuronal
CC alpha-bungarotoxin binding protein (ABBP). They were isolated from
CC an 18-day embryo chick brain cDNA library using a probe (see also
CC T59198) based on the N-terminal amino acid sequence of chicken
CC brain ABBP. The probe isolated partial clone pch29-1, which
CC encoded the N-terminal portion of alpha1. A subclone, pch29-3
CC (ATCC 40611), was used to rescreen the library, yielding clone
CC pch31-1 (ATCC 40640), which encoded the entire alpha2 sequence.
CC A probe based on the C-terminal region of pch31-1 was used to obtain
CC clone pch44-1 (ATCC 40639), encoding the C-terminal portion of
CC alpha1. The cDNA clones can be used as probes to identify further
CC ABBP subunits and in the recombinant prodn. of ABBP.
CC Sequence 2101 BP; 582 A; 406 C; 454 G; 659 T;

Query Match 9.6%; Score 61.6; DB 1; Length 2101;
Best Local Similarity 70.7%; Pred. No. 8.3e-09;
Matches 82; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

OY 526 CTGCGATATTGATGAGCGCTTGACGCCACATTCACACATGCTTGATTC 585
DB 421 CTATACACAGTCTGATGAAAGATTATGCTACATTCACACATTAAGTTTACTCAATTC 480
OY 586 TTCGGGACATTCGACATGACCTGCTCCAGCAGCATATTCAAGAGTCTGCTACATCG 641
DB 481 CTCGTGATCTGTCATATATTCCTCCAGGCAATTTGAAGAGACATGTTACATTG 536

```

```

OS Rattus rattus.
FH Key Location/Qualifiers
FT cds 61..1548
FT *tag= a
FT signal_peptide 61..120
FT *tag= b
FT mat_peptide 121..1545
FT *tag= C
FT *tag= nACHR beta 4 subunit
FT misc_feature 416..417
FT *tag= d
FT label_splice site 1409..1410
FT *tag= e
FT label_splice site 1569..1612
FT *tag= f
FT number= 1
FT 1625..1668
FT *tag= g
FT number= 2
FT 1683..1726
FT *tag= h
FT number= 3
FT 2436..2441
FT *tag= 1
FT conflict
FT *tag= 1
FT *note= conflict between genomic and cDNA sequences"

polya_signal
FT number= 3
FT 2436..2441
FT *tag= 1
FT conflict
FT *tag= 1
FT *note= conflict between genomic and cDNA sequences"

MO9010648-A.
PD 20-SEP-1990.
PF 14-MAR-1990; U01403.
PR 14-MAR-1989; US-321384.
PR 12-MAR-1990; US-492555.
PA (SALK ) SALK INST FOR BIOL. STUD.
PI Heinemann SF; Deneris ES; Duvoisin RM; Patrick JW;
DR WPI: 90-304987/40.
P-PSDB: R07143.
PT New neuronal nicotinic acetylcholine receptor - compns. contg.
PT beta 4 sub-unit and DNA sequences encoding them.
PS Claim 11; Fig 3; 47pp; English.
CC The sequence encodes a novel neuronal nicotinic acetylcholine
CC receptor subunit, beta 4. Plasmid pZPC13 (ATCC 67893) was
CC isolated from a cDNA library in lambda ZAP II prep. from rat
CC mRNA. The sequence of the cDNA in the region encoded by the 5th
CC exon is identical to that determined for the genomic clone with
CC the exception of a sustn. at posn. 720 of a t for a c (this does
CC not alter the AA sequence); this may be due to a polymorphism
CC between the rat strains used for the genomic library and the cDNA
CC library, or could have resulted from a reverse transcriptase error
CC during prepn. of the cDNA. A sequence, the core of which is 44 bp
CC long is repeated three times at the beginning of the 3' UT region.
CC The function of this remains unknown. The protein subunit can
CC combine with the known subunits, alpha 2, 3, and -4, and beta 2
CC to form previously unknown functional receptors. The new subunit
CC is expressed in the central and peripheral nervous systems and in
CC PC12 cells.
CC Sequence 2460 BP; 519 A; 767 C; 587 G; 587 T;

Query Match 7.0%; Score 44.6; DB 1; Length 2460;
Best Local Similarity 61.7%; Pred. No. 0.0012;
Matches 71; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

OY 527 TCGCATATTGCTGATGAGCGCTTTGACGCCACATTCACACATGCTTGATTC 586
DB 409 TACAACAATGCGGATGACGACCATATGAGTGTCTGTACACCAAGTATGTGGTCC 468
OY 587 TCTGGGACATTCGACATGACCTGCTCCAGCAGCATATTCAAGAGTCTGCTACATCG 641
DB 469 AACGGCACATTCGACATGACCTGCTCCAGGCAATTTGAAGAGACATGTTACATTG 523

```

RESULT 7  
X53491 standard; DNA; 114955 BP.  
ID X53491;  
AC X53491;  
DT 05-JUL-1999 (first entry)  
DE Human adenosine A1 receptor antisense oligonucleotide fragment.  
KW Antisense oligonucleotide; multiple target; antisense treatment;  
KW impaired respiration; inflammation; lung disease;  
KW pulmonary vasoconstriction; inflammation; allergic rhinitis;  
KW acute asthma; allergy; asthma; impeded respiration;  
KW respiratory distress syndrome; pain; cystic fibrosis;  
KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;  
KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;  
KW colon cancer; breast cancer; lung cancer; pancreatic cancer;  
KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;  
KW prostate cancer; ss.  
OS Synthetic.  
PN MO913886-A1.  
PD 25-MAR-1999.  
17-SEP-1998; 019419.  
09-JUN-1998; US-093972.  
17-SEP-1997; US-059160.  
(UYEC-) UNIV EAST CAROLINA.  
PI NYCE JM;  
PT MPI; 99-229400/19.  
PT New antisense oligonucleotides used in treatment of, e.g. pulmonary  
PT vasoconstriction  
PS Disclosure; Page 37; 120pp; English.  
CC The specification describes antisense oligonucleotides (X52869-X55271)  
CC directed against at least 2 mRNAs selected from target genes, coding and  
CC non-coding regions of RNAs corresponding to target genes, gene  
CC initiation codons, genomic flanking regions, intron-exon borders, the  
CC 5'-end, the 3'-end and the junction between coding and non-coding  
CC regions and all segments of RNAs encoding proteins associated with one  
CC or more diseases, conditions or mixtures. The antisense oligonucleotides  
CC may be derived from sequences X55272-74. These multiple target  
CC oligonucleotides (specifically X55180-271) can be used for the antisense  
CC treatment of diseases and conditions. Typical diseases and conditions  
CC are those associated with impaired respiration and inflammation,  
CC including lung diseases, pulmonary vasoconstriction, inflammation,  
CC allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,  
CC respiratory distress syndrome, pain, cystic fibrosis, pulmonary  
CC hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive  
CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,  
CC carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic  
CC cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic  
CC metastases, as well as all types of cancers which may metastasize or have  
CC metastasized to the lungs, including breast and prostate cancer.  
CC Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T;

Query Match 6.7%; Score 43; DB 1; Length 114955;  
Best Local Similarity 41.2%; Pred. No. 0.026;  
Matches 73; Conservative 15; Mismatches 89; Indels 0; Gaps 0;

QY 7 GCCACATAGTCCCGCCCACTCTGCTGCTCCCTTCATTTCCAGCCGCTCCACG 66  
||| :  
Db 105096 GCCGNNNDNCCBGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105155  
||| :  
QY 67 AGGTACAGCGCGCGCGGAGAGTGTGAGCGCGAGAGCTCGCGCGCGCGCGCTGT 126  
||| :  
Db 105156 CGGGCCGCGNNNDNBBGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105215  
||| :  
QY 127 GCGCGCGCGCGCATGACAGCGCTCGGAGTGTCTTTTCGCGCGCTCCCGCG 183  
||| :  
Db 105216 GGGCGCGNNNDNNGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105272  
||| :

RESULT 8  
T95785 standard; cDNA to mRNA; 2383 BP.  
ID T95785;  
AC T95785;  
DT 22-MAY-1998 (first entry)

DE Maize starch synthase type I cDNA  
KW Maize; starch synthase type I; starch; ds.  
OS Zea mays.  
FH Key  
FT CDS  
FT Location/Qualifiers  
FT 2..1951  
FT /\*tag= a  
FT /\*product= starch\_synthase\_type\_I  
DE19619918-A1.  
PD 20-NOV-1997.  
PF 17-MAY-1996; 019918.  
PR 17-MAY-1996; DE-019918.  
PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.  
PI Froberg C, Kossman J;  
DR MPI; 98-000821/01.  
P-PDB: W38218.  
PT DNA encoding maize starch synthase type I protein - for producing  
PT transgenic plants  
PS Claim 1; Pages 16-20; 23pp; German.  
CC The present sequence encodes maize starch synthase type I, useful  
CC in the production of starch. Starch can be used in various  
CC conventional starch applications, e.g. starch hydrolyzate products,  
CC foods, papermaking, adhesives, textiles, building materials, soil  
CC stabilization, agrochemicals, pharmaceuticals, cosmetics, coal  
CC briquettes, ore and coal slurries, foundry casting, rubber, leather  
CC and synthetic polymers. The enzyme produces a starch started to have  
CC different physicochemical properties, especially viscosity and  
CC gelling properties, from wild type starch.  
SQ Sequence 2383 BP; 582 A; 546 C; 633 G; 622 T;

Query Match 6.7%; Score 42.8; DB 1; Length 2383;  
Best Local Similarity 55.3%; Pred. No. 0.0039;  
Matches 83; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 58 GCTCCAGAGGGTACAGCGCGCGGAGAGTGTGAGCGCGGAGAGCTCGCGCGGCC 117  
||| :  
Db 121 GCTCCAGCGCGGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 180  
||| :  
QY 118 CGGCTGGGCGCGCGCGCATACAGCGGCTCGGAGTGTCTTTTCGCGCGCCCTCC 177  
||| :  
Db 181 GCGCCGATGACACCGCGCGCTCTGCGCGCGCGCGCGCGCGCGCTCTCTGCGCGCC 240  
||| :  
QY 178 CGCGAGGTGAGGAGATGTCATGTC 207  
||| :  
Db 241 GCGCGAGCCACGAGGTGAGCGCGCATGTC 270  
||| :

RESULT 9  
X28103/c  
ID X28103 standard; DNA; 2106 BP.  
AC X28103;  
DT 11-JUN-1999 (first entry)  
DE Freac11 gene.  
KW Freac11; transcription factor; adipose tissue expressed gene; obesity;  
KW lipid metabolism; adipocyte differentiation; obesity-related condition;  
KW non-insulin dependent diabetes mellitus; cardiovascular disease;  
KW catabolic condition; anorexia; bulimia; therapy; ss.  
OS Homo sapiens.  
PN MO9854216-A1.  
PD 03-DEC-1998.  
PF 26-MAY-1998; SE0989.  
PR 26-MAY-1997; SE-001963.  
PA (PMA) PHARMACIA & UPJOHN AB.  
PI Carlsson P, Enerbaeck S;  
DR MPI; 98-070211/06.  
P-PDB: Y01097.  
PT New transcription factor designated FREAC11 - which regulates  
PT adipose tissue expressed genes involved in lipid metabolism and  
PT adipocyte differentiation, used to, e.g. inhibit gene expression  
PS Claim 1; Fig 1; 23pp; English.  
CC This sequence represents the Freac11 gene of the invention. The encoded  
CC protein has transcriptional regulative function directed against adipose  
CC tissue expressed genes. The protein is involved in lipid metabolism



```

FT cds 87..1583
PN W09641876-A1. /tag- a
PD 27-DEC-1996.
PF 07-JUN-1996; U09775.
PR 07-JUN-1995; US-484722.
PA (SIBI-) SIBIA NEUROSCIENCES INC.
PI Elliott KJ, Harpold MM;
DR WPI: 97-065463/06.
P-PSDB: W09027.
PT Nucleic acids encoding nicotinic acetylcholine receptor sub-units -
PS used in screening to determine the effect of drugs on the receptor
CC disclosure: Page 83-85; 108pp; English.
CC A DNA sequence (T48241) codes for the beta-4 subunit (W09027) of
CC the human neuronal nicotinic acetylcholine receptor (nAChR). Host
CC cells, esp. mammalian cells or amphibian oocytes, carrying beta-4
CC nucleic acids, opt. in combination with other alpha and/or beta
CC subunit nucleic acids (see also T48232-40), express recombinant
CC nAChR subunits useful for identifying cpds. that modulate the
CC activity of human nAChRs.
Sequence 1915 BP; 374 A; 631 C; 505 G; 405 T;

Query Match 5.7%; Score 36.6; DB 1; Length 1915;
Best Local Similarity 57.4%; Pred. No. 0.26; Mismatches 49; Indels 0; Gaps 0;
Matches 66; Conservative 0;

QY 527 TGGCATTTGCTGATGAGCGCTTGGACGACATTCACACTACGTTGGTAATCT 586
DB 438 TACACACAGCGCCGACGGAGCTATGAGTGTCTGTCTACACCACTGATAGCGGTC 497
QY 587 TCTGGCGATTGCCAGTACCTGCTCCAGGCATATTCAGAGTTCTCTGCTACATCG 641
DB 498 AACGCGAGCGTCTCTGTGCTGCCCTCCGATCTACAGAGCGCCTCAAGATTG 552

RESULT 13
V68408/c
ID V68408 standard; cDNA to mRNA; 5561 BP.
DR 05-MAY-1999 (first entry)
DE Human BAZ1-beta cDNA #1
KW Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;
KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
OS Homo sapiens.
FH Key Location/Qualifiers
CDS 346..4929
/Tag- a "BAZ1-beta"
/product- "BAZ1-beta"
PN W09847920-A1.
PD 29-OCT-1998.
PF 17-APR-1998; J01783.
PR 24-OCT-1997; JP-310027.
PR 18-APR-1997; JP-116570.
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PI Jones MH;
DR WPI: 98-583603/49.
P-PSDB: W81172.
PT Transcriptional regulator gene family containing bromodomain - may
PT be expressed in testis tissue and is useful for treatment of cancer
PT and other proliferative disorders
PS Claim 2; Page 125-137; 187pp; Japanese.
CC This sequence encodes the human BAZ1-beta protein, a member of a
CC family of transcriptional regulator genes containing a bromodomain (BAZ,
CC Bromodomain with Atypical zinc finger) which are expressed specifically
CC in testis tissue and also in certain tumour lines. Transgenic cells may
CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
CC and other proliferative disorders, and in screening of compounds for
CC their binding ability to the expression products (e.g. for use as drugs
CC by modulation of transcriptional regulation).
Sequence 5561 BP; 1665 A; 1261 C; 1471 G; 1164 T;

```

```

Query Match 5.7%; Score 36.4; DB 1; Length 5561;
Best Local Similarity 58.2%; Pred. No. 0.51; Mismatches 46; Indels 0; Gaps 0;
Matches 64; Conservative 0;

QY 6 CGCCACATAGCTCCCGCCAGAGTCCTGCGCCCTTGCATTTTCACAGCCGCTCCAC 65
DB 156 CGGCCAGAGCTCCCGCCACACCGCCGCTCCAGACAGCCCCCGCCGACCTCCCT 97
QY 66 GAGGCTCAGCGCGCGGAGAGAGTGAGCCGCGAGAGCTGCGCGGAGG 115
DB 96 TCGGCTCCCGCGCGCGGACGACGACTCTCTCAGCAGCAGCAGCGAGG 47

RESULT 14
V68409/c
ID V68409 standard; cDNA to mRNA; 5573 BP.
DR 05-MAY-1999 (first entry)
DE Human BAZ1-beta cDNA #2.
KW Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;
KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
KW BAZ2-beta; treatment; cancer; proliferative disorder; screening; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
CDS 346..4941
/Tag- a
/product- "BAZ1-beta"
PN W09847920-A1.
PD 29-OCT-1998.
PF 17-APR-1998; J01783.
PR 24-OCT-1997; JP-310027.
PR 18-APR-1997; JP-116570.
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PI Jones MH;
DR WPI: 98-583603/49.
P-PSDB: W81173.
PT Transcriptional regulator gene family containing bromodomain - may
PT be expressed in testis tissue and is useful for treatment of cancer
PT and other proliferative disorders
PS Claim 2; Page 145-157; 187pp; Japanese.
CC This sequence encodes the human BAZ1-beta protein, a member of a
CC family of transcriptional regulator genes containing a bromodomain (BAZ,
CC Bromodomain with Atypical zinc finger) which are expressed specifically
CC in testis tissue and also in certain tumour lines. Transgenic cells may
CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
CC and other proliferative disorders, and in screening of compounds for
CC their binding ability to the expression products (e.g. for use as drugs
CC by modulation of transcriptional regulation).
Sequence 5573 BP; 1668 A; 1267 C; 1472 G; 1166 T;

Query Match 5.7%; Score 36.4; DB 1; Length 5573;
Best Local Similarity 58.2%; Pred. No. 0.52; Mismatches 46; Indels 0; Gaps 0;
Matches 64; Conservative 0;

QY 6 CGCCACATAGCTCCCGCCAGAGTCCTGCGCCCTTGCATTTTCACAGCCGCTCCAC 65
DB 156 CGGCCAGAGCTCCCGCCACACCGCCGCTCCAGACAGCCCCCGCCGACCTCCCT 97
QY 66 GAGGCTCAGCGCGCGGAGAGAGTGAGCCGCGAGAGCTGCGCGGAGG 115
DB 96 TCGGCTCCCGCGCGCGGACGACGACTCTCTCAGCAGCAGCAGCGAGG 47

RESULT 15
O84888/c
ID O84888 standard; cDNA to mRNA; 1942 BP.
AC O84888;
DR 26-OCT-1995 (first entry)
DE DNA encoding cytoplasmic tyrosine kinase.

```



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:05:19 ; Search time 1962.17 Seconds  
(without alignments)  
1038.941 Million cell updates/sec

Title: US-08-956-518A-103

Perfect score: 641  
Sequence: 1 CAGCCGCCACATAGCTCCC.....CAGAGTTCCTGCTACATCG 641

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

ibase :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_com:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl1:\*  
10: gb\_pl2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_st:\*  
14: gb\_st:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_v1:\*  
18: em\_fun:\*  
19: em\_htg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_om:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_ro:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_v1:\*  
34: gb\_htg1:\*  
35: gb\_htg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_pat:\*  
39: em\_pat:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pr4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	641	100.0	641	11	AF029838	AF029838 Homo sapi

2	503	78.5	689	11	AF029839	AF029839 Homo sapi
3	217.2	33.9	291	9	HS2E10F	Z58126 H.sapiens C
4	196.6	30.7	137699	11	AC005630	AC005630 Homo sapi
5	194.2	30.3	298	9	HS2G7F	Z60646 H.sapiens C
6	192	30.0	1712	11	AF036903	AF036903 Homo sapi
7	171.8	26.8	192439	35	AC007562	AC007562 Homo sapi
8	163.4	25.5	178307	35	AC006328	AC006328 Homo sapi
9	156.8	24.5	1836	11	AF037646	AF037646 Homo sapi
10	108	16.8	1559	9	HUMA7A	L53827 Human a7 n1
11	108	16.8	1559	10	HSARA7A	X70297 H.sapiens m
12	108	16.8	1509	10	HSNAC7A	Y08420 H.sapiens m
13	108	16.8	1876	10	HSN62436	U62436 Human nicot
14	106.4	16.6	1555	10	HSCHRN7A	U23141 H.sapiens C
15	106.4	16.6	1977	10	HSU04583	U40583 Human alpha
16	92.6	14.4	5281	9	HSU08191	U08191 Human R Kap
17	92	14.4	1848	12	MUSNARS	L37663 Mus musculi
18	90.4	14.1	2106	12	RATNARAD	L31619 Rattus ratt
19	90.4	14.1	3030	12	S53987	S53987 nicotinic r
20	88.8	13.9	1551	3	BR7A7	X93604 B.taurus nr
21	74.4	11.6	2696	4	GGATNAREC	X52295 Chicken alp
22	74.4	11.6	2037	4	GGNNA7A7A	X68586 G.gallus mr
23	62.4	9.7	171	9	HS165HIF	Z57156 H.sapiens C
24	61.8	9.6	180	9	HS115G2F	Z54415 H.sapiens C
25	61.6	9.6	2090	4	GGARNAREC	X52296 Chicken alp
26	60.4	9.4	189	9	HS31DBR	Z61332 H.sapiens C
27	58.2	9.1	3029	37	AF143847	AF143847 Heliothis
28	50	7.8	89871	37	AC005890	AC005890 Drosophila
29	48.2	7.5	32039	1	SC7A1	AL034447 Streptomy
30	46.4	7.2	336	14	G47584	G47584 225357.1 2e
31	46.2	7.2	76947	36	AC004326	AC004326 Drosophila
32	46.2	7.2	3629	37	AF143846	AF143846 Heliothis
33	44.6	7.0	2493	12	RNNIACETE	X15834 R.norvegicus
34	44.6	7.0	2461	12	RNU42976	U42976 Rattus norv
35	44	6.9	1008	12	RATNACHRFS	M33952 Rat neurona
36	42.8	6.7	143664	35	AC007291	AC007291 Drosophila
37	42.4	6.6	3459	10	HSAR12582	AJ012582 Homo sapi
38	42.4	6.6	785	10	HSAR13727	AJ13727 Homo sapi
39	42.2	6.6	107889	11	AC004519	AC004519 Homo sapi
40	41.8	6.5	1574	11	AF069682	AF069682 Homo sapi
41	41.6	6.5	120766	42	AC004150	AC004150 Homo sapi
42	41.2	6.4	2672	3	RABTRHP	T05603 Rabbit mult
43	41.2	6.4	4085	4	CHKACHR02	K02904 Chicken nlc
44	41.2	6.4	3289	10	HSMEF1	Y08223 H.sapiens M
45	41.2	6.4	173422	35	AC007041	AC007041 Homo sapi

## ALIGNMENTS

RESULT 1  
AF029838 641 bp mRNA 16-DEC-1998  
DEFINITION Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.  
ACCESSION AF029838  
NID 93757793  
VERSION AF029838.1 GI:3757793  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 641)  
Gault,J., Robinson,M., Berger,R., Drebing,C., Logel,J., Hopkins,J.,  
Moore,T., Jacobs,S., Meriwether,J., Choi,M.J., Kim,E.J., Walton,K.,  
Buiting,K., Davis,A., Breese,C., Freedman,R. and Leonard,S.,  
Genomic organization and partial duplication of the human alpha7  
neuronal nicotinic acetylcholine receptor gene  
Genomics 52 (2), 173-185 (1998)  
JOURNAL 99000837  
MEDLINE  
REFERENCE  
2 (bases 1 to 641)  
Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R.,  
Breese,C., Davis,A., Hopkins,J. and Freedman,R.  
Direct Submission  
TITLE





QY 541 TGAGCGCTTGAGCCACATTCACACTAGCGTGTGTAATTCCTTCGGGATTCGCA 600  
|||||  
Db 589 TGAGCGCTTGAGCCACATTCACACTAGCGTGTGTAATTCCTTCGGGATTCGCA 648  
QY 601 GTACTGCTCCAGCATATTCAGAGTTCTCTCATCG 641  
|||||  
Db 649 GTACTGCTCCAGCATATTCAGAGTTCTCTCATCG 689  
RESULT 3  
HS2E10F 291 bp DNA PRI 18-OCT-1995  
LOCUS H.sapiens Cpg Island DNA genomic MseI fragment, clone 2e10, forward  
DEFINITION read cpg2e10.fttld.  
ACCESSION 258126  
NID 91029357  
VERSION 258126.1 GI:1029357  
KEYWORDS Cpg Island; genomic MseI fragment.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 291)  
AUTHORS MacDonald, M., Huckle, E., Wilkinson, P. and Mickle, G.  
TITLE Direct Submission  
JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1HQ, England. E-mail contact: humuery@sanger.ac.uk  
COMMENT 2 (bases 1 to 291)  
Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.  
Purification of Cpg Islands using a methylated DNA binding column  
Nat. Genet. 6 (3), 236-244 (1994)  
94282070  
Vector: pGEM-52f(-)  
Clones are available from the UK MRC Human Genome Mapping Project  
Resource Centre, Hinxton, Cambridgeshire CB10 1HQ, UK. See URL:  
http://www.hmp.mrc.ac.uk/ for details  
or contact: biohelp@hmp.mrc.ac.uk.  
FEATURES  
source Location/Qualifiers  
1..291  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/sex="male"  
/dev\_stage="adult"  
/tissue\_type="blood"  
/clone\_id="CGI-1"  
/clone="2e10"  
BASE COUNT 44 a 95 c 103 g 45 t 4 others  
TGIN  
Query Match 33.9%; Score 217.2; DB 9; Length 291;  
Best Local Similarity 97.3%; Pred. No. 3.5e-45;  
Matches 219; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 1 CAGGCGCCACATAGCTCCCGCAATCTCCGCTGCCCTTCGATTTCCAGCCGCGT 60  
|||||  
Db 67 CAGGCGCCACATAGCTCCCGCAATCTCCGCTGCCCTTCGATTTCCAGCCGCGT 126  
QY 61 CCCACAGAGGTCAAGCGCGGAGAGGTGAGCCGAGAGCTGGCGGGGGCCCG 120  
|||||  
Db 127 CCCACAGAGGTCAAGCGCGGAGAGGTGAGCCGAGAGCTGGCGGGGGCCCG 186  
QY 121 CCTGTGGCGCGCCATGACAGCGGCTCGGACTGCTCTTTCCGCGCCCTCCCG 180  
|||||  
Db 187 CCGGTGTGTCGCGGATGACAGCGGCTCGGAGAGCTCTTTCCGCGCCCTCCCG 246  
QY 181 CCGAGGTGAGGAGAGATGCTTCATGAGGTTCAAGGCCAAAC 225  
|||||  
Db 247 CCGAGGTGAGGAGAGATGCTTCATGAGGTTCAAGGCCAAAC 291  
RESULT 4

AC005630  
LOCUS AC005630 137699 bp DNA PRI 15-JAN-1999  
DEFINITION Homo sapiens clone DJ1129D05, complete sequence.  
ACCESSION AC005630  
NID 94159882  
VERSION AC005630.1 GI:4159882  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 137699)  
AUTHORS Waterston, R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
2 (bases 1 to 137699)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (08-SEP-1998) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
3 (bases 1 to 137699)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (15-JAN-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Jan 15, 1999 this sequence version replaced gi:3550965.  
COMMENT Location/Qualifiers  
FEATURES  
source 1..137699  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DJ1129D05"  
BASE COUNT 36908 a 30670 c 32412 g 37709 t  
ORIGIN  
Query Match 30.7%; Score 196.6; DB 11; Length 137699;  
Best Local Similarity 80.2%; Pred. No. 1.2e-39;  
Matches 243; Conservative 0; Mismatches 59; Indels 1; Gaps 1;  
QY 1 CAGGCGCCACATAGCTCCCGCAATCTCCGCTGCCCTTCGATTTCCAGCCGCGC 59  
|||||  
Db 47458 CAGGCGCCACATAGCTCCCGCAATCTCCGCTGCCCTTCGATTTCCAGCCGCGC 47517  
QY 60 TCCACAGAGGTCAAGCGCGGAGAGGTGAGCCGAGAGCTGGCGGGGGCCCG 119  
|||||  
Db 47518 TCCACAGAGGTCAAGCGCGGAGAGGTGAGCCGAGAGCTGGCGGGGGCCCG 47577  
QY 120 GCCTGTGGCGCGCCATGACAGCGGCTCGGACTGCTCTTTCCGCGCCCTCCCG 179  
|||||  
Db 47578 GCCTGTGGCGCGCCATGACAGCGGCTCGGACTGCTCTTTCCGCGCCCTCCCG 47637  
QY 180 CCGAGGTGAGGAGAGATGCTTCATGAGGTTCAAGGCCAAACGAGTACAGCGCT 239  
|||||  
Db 47638 CCGAGGTGAGGAGAGATGCTTCATGAGGTTCAAGGCCAAACGAGTACAGCGCT 47697  
QY 240 CTATCTTCAGAGAGAACAGAGACACAGCCGCTCAGCCGACCAACATTAGA 299  
|||||  
Db 47698 CTATCTTCAGAGAGAACAGAGACACAGCCGCTCAGCCGACCAACATTAGA 47757  
QY 300 TTA 302  
Db 47758 TTA 47760  
RESULT 5  
HS2G7F 298 bp DNA PRI 19-OCT-1995  
LOCUS H.sapiens Cpg Island DNA genomic MseI fragment, clone 2g7, forward  
DEFINITION read cpg2g7.fta.  
ACCESSION Z60646  
NID 91032750

VERSION 260646.1 GI:1032750  
 KEYWORDS CPG island; genomic MseI fragment.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 298)  
 TITLE Macdonald, M., Huckle, E., Wilkinson, P. and Micklem, G.  
 JOURNAL Direct Submission  
 Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk  
 2 (bases 1 to 298)  
 REFERENCE Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.  
 AUTHORS Purification of CPG islands using a methylated DNA binding column  
 JOURNAL Nat. Genet. 6 (3), 236-244 (1994)  
 MEDLINE 94282070  
 COMMENT Vector: pGEM-5zf(-)  
 Clones are available from the UK MRC Human Genome Mapping Project  
 Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:  
 http://www.hgmp.mrc.ac.uk/ for details  
 or contact: biohelp@hgmp.mrc.ac.uk.  
 Location/Qualifiers  
 1. 298  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /sex="male"  
 /dev\_stage="adult"  
 /tissue\_type="blood"  
 /clone\_lib="CGI-1"  
 /clone="297"

BASE COUNT 47 a 99 c 102 g 48 t 2 others  
 ORIGIN

Query Match 30.3%; Score 194.2; DB 9; Length 298;  
 Best Local Similarity 94.5%; Pred. No. 2,3e-39;  
 Matches 223; Conservative 0; Mismatches 9; Indels 4; Gaps 2;

QY 4 GCCGCACATAGTCTCCGCGCAAGTCTCGGTGCGCCCTTGCCATTTTCACGCGCGCTCC 63  
 DB 67 GCCGCACATAGTCTCCGCGCAAGTCTCGGTGCGCCCTTGCCATTTTCACGCGCGCTCC 126

QY 64 ACAGAGTCAAGCGCGCGGAGAGAGTGGAGCGCGGAGAGCTGGCGCGGCGCGCGCT 123  
 DB 127 ACAGAGTCAAGCGCGCGGAGAGAGTGGAGCGCGGAGAGCTGGCGCGGCGCGCT 183

QY 124 GGTGGCGCGCGCATGACAGAGCGGCGGAGTGGCTCTTTCCGCGCGCGCTCCGCGG 183  
 DB 184 GGTGGCGCGCGCATGACAGAGCGGCGGAGTGGCTCTTTCCGCGCGCGCTCCGCGG 243

QY 184 AGGTGAGGAGAGATGTCATGTCAGAGGTTCAGAGCCCAACCGAAGTACTGGCCT 239  
 DB 244 A-GTAGAGGGAGATGTCATGTCAGAGGTTCAGAGCCCAACCGAAGTACTGGCCT 298

RESULT 6  
 AF036903 1712 bp mRNA PRI 16-OCT-1998  
 LOCUS Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor  
 DEFINITION AF036903  
 ACCESSION AF036903  
 NID 93757807  
 VERSION 3757807  
 KEYWORDS AF036903.1 GI:3757807  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 1712)  
 REFERENCE Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J.,  
 Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K.,  
 Bulting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.  
 COMMENT Genomic organization and partial duplication of the human alpha7

JOURNAL neuronal nicotinic acetylcholine receptor gene (CHRNA7)  
 MEDLINE Genomics 52 (2), 173-185 (1998)  
 REFERENCE 99000837  
 AUTHORS 2 (bases 1 to 1712)  
 TITLE Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,  
 Breese, C., Davis, A., Hopkins, J. and Freedman, R.  
 JOURNAL Direct Submission  
 Submitted (03-DEC-1997) Psychiatry, University of Colorado Health  
 Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA  
 1. 1712  
 Location/Qualifiers  
 1. 1712  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /db\_xref="dbEST:W52861"  
 /db\_xref="dbEST:AA037389"  
 /chromosome="15"  
 /clone="321242"  
 /clone\_lib="Soares papathroid tumor NBHRA library,  
 Reseach Genetics/IMAGE Consortium, LBNL"  
 <1. 1712  
 /note="alpha-7 neuronal nicotinic acetylcholine receptor;  
 alternatively spliced"

BASE COUNT 356 a 489 c 457 g 410 t  
 ORIGIN

Query Match 30.0%; Score 192; DB 11; Length 1712;  
 Best Local Similarity 100.0%; Pred. No. 1e-38;  
 Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 ATTGATGCAAAATATTGCAATCTACAGCAATTTTCAGTTCGAATGCTAATCCAGCAAT 509  
 DB 10 ATTGATGCAAAATATTGCAATCTACAGCAATTTTCAGTTCGAATGCTAATCCAGCAAT 69

QY 510 TGTGATAGCTCAAAACGCGATATGCTGATGAGCGCTTACGCCCATTCACACACTA 569  
 DB 70 TGTGATAGCTCAAAACGCGATATGCTGATGAGCGCTTACGCCCATTCACACACTA 129

QY 570 ACGTGTGTGTAATCTTCTGCGCATTCGCGCATGCTGCTCCAGGCAATATCAAGAGTT 629  
 DB 130 ACGTGTGTGTAATCTTCTGCGCATTCGCGCATGCTGCTCCAGGCAATATCAAGAGTT 189

QY 630 CCTGCTACATCG 641  
 DB 190 CCTGCTACATCG 201

RESULT 7  
 AC007562 192439 bp DNA HTG 05-JUN-1999  
 LOCUS Homo sapiens clone NH0497C14, WORKING DRAFT SEQUENCE, 4 unordered  
 DEFINITION AC007562  
 ACCESSION AC007562  
 NID 95001485  
 VERSION AC007562.2 GI:5001485  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 192439)  
 REFERENCE Waterston, R.H.  
 AUTHORS The sequence of Homo sapiens clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 192439)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-MAY-1999) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 ON Jun 5, 1999 this sequence version replaced g1:4827309.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. The true order of the pieces



MEDLINE	99000837
REFERENCE	2 (bases 1 to 1836)
AUTHORS	Leonard,S., Gaul,J., Logel,J., Drebing,C., Robinson,M., Berger,R. Brees,E., Davis,A., Hopkins,V. and Freedman,R.
TITLE	Direct Submission
JOURNAL	Submitted (03-DEC-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
FEATURES	Location/Qualifiers
SOURCE	1..1836
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/db_xref="dbEST:W03952"
	/db_xref="dbEST:N73891"
	/chromosome="15"
	/clone="297134"
	/clone.lib="Soares fetal liver spleen library, Research Genetics/IMAGE Consortium, LLNL"
	/note="unprocessed mRNA with intron"
	<1..1836
	/note="alpha-7 neuronal nicotinic acetylcholine receptor precursor; Intron included in 5' region"
BASE COUNT	383 a 503 c 469 g 481 t
ORIGIN	
Query Match	24.5%; Score 156.8; DB 11; Length 1836;
Best Local Similarity	98.8%; Pred. No. 7.9e-30;
Matches 158; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
OY	482 TTTCAGTCCATTGCTATCCAGATTGTGGATGCTGCACAACTGCATATGCTGAT 541
DB	164 TTCTAGTTCCATTGCTATCCAGCATTTGTGAATGCGCAACTGCATATGCTGAT 223
OY	542 GAGCCTTGACGCCACATTCACACTAAGCTGTGGTAATCTCTGGCATTCSCAG 601
DB	224 GAGCGCTTGACGCCACATTCACACTAAGCTGTGGTAATCTCTGGCATTCSCAG 283
OY	602 TACCTGCTCCAGCATATTCAGAAGTCTCTGCTACATCG 641
DB	284 TACCTGCTCCAGCATATTCAGAAGTCTCTGCTACATCG 323
RESULT 10	
HUMANR	
LOCUS	HUMA7NRK 1555 bp mRNA PRI 18-AUG-1994
DEFINITION	Human a7 nicotinic acetylcholine receptor mRNA.
SSION	L25827 L25827 9438616
VERSION	L25827.1 GI:438616
KEYWORDS	nicotinic acetylcholine receptor alpha 7 subunit; transmembrane protein.
SOURCE	Homo sapiens (Library: Clontech HL1065b; ATCC 37433) fetus brain CDNA to mRNA.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	Dovelette-Stamm,L., Monteggia,L.M., Donnelly-Roberts,D., Wang,M.T., Lee,J., Tian,J. and Giordano,T.
TITLE	Cloning and sequence of the human a7 nicotinic acetylcholine receptor
JOURNAL	Drug Dev. Res. (1993) in press
FEATURES	Location/Qualifiers
SOURCE	1..1559
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/dev_stage="fetus"
	/tissue_type="brain"
	/tissue_lib="Clontech HL1065b; ATCC 37433"
BASE COUNT	314 a 463 c 440 g 342 t
ORIGIN	

Query Match 16.8% Score 108; DB 9; Length 1559;  
Best Local Similarity 95.7%; Pred. No. 1.7e-17;  
Matches 111; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 526 CTGGCATATTCGTAGTGAAGCGCTTTGACGCCACATATTCACACTAAGCTGTGTAATTC 585  
||| |  
Db 319 CTAATAACAGTGCTGTAGAGCCGCTTTGAGGCCACATATTCACACTAAGCTGTGTAATTC 378

QY 586 TTTCGGGCATTGCCAGTACCTGCTCCTCCAGGCAATATTCAGAAGTCTCTCATATCG 641  
|||||  
Db 379 TTTCGGGCATTGCCAGTACCTGCTCCTCCAGGCAATATTCAGAAGTCTCTCATATCG 434

RESULT 11

HSARA7A LOCUS HSARA7A 2087 bp mRNA PRI 01-JUN-1994

DEFINITION H.sapiens mRNA for neuronal nicotinic acetylcholine receptor alpha-7 subunit.

ACCESSION X70297

NID 9496606

VERSION X70297.1 GI:496606

KEYWORDS neuronal nicotinic acetylcholine receptor alpha-7 subunit.  
human.

ORGANISM Homo sapiens  
Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 2087)  
Peng,X., Katz,M., Gerzanich,V., Anand,R. and Lindstrom,J.  
Human alpha 7 acetylcholine receptor: cloning of the alpha 7  
subunit from the SH-SY5Y cell line and determination of  
pharmacological properties of native receptors and functional alpha  
7 homomers expressed in Xenopus oocytes  
Mol. Pharmacol. 45 (3), 546-554 (1994)

JOURNAL MEDLINE 94195283

REFERENCE 2 (bases 1 to 2087)  
Katz,M.

AUTHORS Direct Submission

TITLE Submitted (04-FEB-1993) M. Katz, Univ of Pennsylvania School of  
Medicine, Dept of Neuroscience, 36th & Hamilton walk, Rm 335  
Stemmer Hall, Philadelphia, PA 19104, USA

JOURNAL location/Qualifiers

FEATURES SOURCE

1. 2087  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="brain"  
/cell\_type="neuroblastoma"  
/cell\_line="SHSY-5Y"  
/clone\_idb="lambda zap II"  
/clone="SHSY3"  
104..175  
/product="neuronal nicotinic acetylcholine receptor  
alpha-7 subunit"  
/codon\_start=1  
/product="neuronal nicotinic acetylcholine receptor  
alpha-7 subunit"  
/protein\_id="CAA49778.1"  
/db\_xref="PID:9496607"  
/db\_xref="GI:496607"  
/db\_xref="SWISS-PROT:P36544"  
/translation="NRCSPGVWLGLAAALSHVSLQGEPORLIKELYNYPNLEPRV  
ANDSQLTYFESLILQIMDVDERKNQVLNTNLQMSMTDHPLOAMSEVPGRKYTF  
PDSQITRPDLILNSADERDATFRHTNYLVNNSGHCQYLDPGIFKSCTIDYRPFPPD  
VDGCKTKFSWSIGWSDLIQMEADISGITPNGEWLDVGIPKRSEERFECKEPPP  
DYFTVTNRRRTLYLGILNLPICVLISALTIPADSGEKISLGIYLLSLTYFM  
LLVAELIMPATSDVPLIAQYFASTMITIGSVVVTVIYQHHNDPDGSKPMKTRVI  
LTLMWCFLMLMKRPEDGVAPACOKHORCSLASVESNAVGPAPSNGNLIIYGPNL  
DYHCVPTDSDGVGVGARMACSPHNDEHLHGQOPREGDPDLAKILLEVRVIANRFQC  
DSEBANVCSEKRFACAYVDRLCLMAFSVFITICTIGILMSAPNFVEAVSKDFA"  
176..1603

mat\_peptide 441 a 601 c 573 g 472 t

BASE COUNT 441 a 601 c 573 g 472 t

RRIGIN



3'UTR 1582 1876 531 g 423 t  
BASE COUNT 369 a 553 c  
ORIGIN

Query Match 16.8%; Score 108; DB 10; Length 1876;  
Best Local Similarity 95.7%; Pred. No. 1.7e-17;  
Matches 110; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 526 CTGCGATATTCCTGATGAGCGCTTTGACGCCACATTCACACTACGTTGGTGAATTC 585  
|||  
Db 414 CTAAACAGTGCCTATGAGCGCTTTGAGCCACATTCACACTACGTTGGTGAATTC 473  
|||  
QY 586 TTCTGGGCGATTGCCAGTACCTGCTCCAGGCATATTCAGAGTCTGCTACATCG 641  
|||  
Db 474 TTCTGGGCGATTGCCAGTACCTGCTCCAGGCATATTCAGAGTCTGCTACATCG 529

LOCUS 1582 bp mRNA PRI 03-MAR-1994  
DEFINITION H.sapiens CHRNA7 mRNA, 3' end.  
ACCESSION 223141  
NID 6457736  
VERSION 223141.1 GI:457736  
KEYWORDS alpha/nicotinic receptor subunit; cholinergic receptor; CHRNA7 gene.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1555)  
AUTHORS Chini, B., Raimond, E., Eljoughen, A.B., Moralli, D., Balzaretto, M. and Heinemann, S.  
TITLE Molecular cloning and chromosomal localization of the human alpha 7-nicotinic receptor subunit gene (CHRNA7)  
JOURNAL Genomics 19 (2), 379-381 (1994)  
MEDLINE 94245214  
AUTHORS Chini, B.  
TITLE Direct Submission  
JOURNAL Submitted (29-JUN-1993) BICE CHINI, CCIPE, Rue de la Cardonille, Montpellier, 34094 Cedex 5, France

FEATURES  
source  
1. 1555  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="retina"  
/clone\_lib="lambda gt11"  
/gene="CHRNA7"  
/product="cholinergic receptor, nicotinic, alpha polypeptide 7"  
/protein\_id="CA80672.1"  
/db\_xref="PIR:G457737"  
/db\_xref="GI:457737"  
/translation="EFQRLKELVKNPLERPVANDSOPLVFSLIOLIMVDE  
KNQVLTINIMLQMSWTDLQIMVSEYPGVKTIVRPDGOIMKPDILLYNSADERPAT  
FHTNVLVNSGHCQYLPQIFKSCYIDRWPFQVQHCRIKGSWSYSGMSLDLQMO  
EADISGYNIGENDLVIGIKRSERYECCKEPYPTVTYRRRTLYGLNLIPIC  
VLISALAVFLPADSGERTIGITVLISLVFMILVAIMPATSDVPLIAQFAS  
TMTISGLSVVTVIVLQHHHDPDGMKMTFVILLNMCAMFLRMKRGEDKVRPAC  
OHKORSSLASVMSAVAPPASNGNLLYIGFGDGVHCVPTPDGGVCGAMPAPT  
HDEHLHGQPPGEGDPLAKILEEVRVIANRFRCQDESAVCSSEMKFACVVDRLCLM  
AFSEVFTICTIGILMSAPNFEAVSKDFA"

mat\_peptide  
gene  
CDS  
3'UTR  
sig\_peptide  
CDS

BASE COUNT 320 a 454 c 434 g 347 t  
ORIGIN

Query Match 16.6%; Score 106.4; DB 10; Length 1555;  
Best Local Similarity 94.8%; Pred. No. 4.2e-17;  
Matches 110; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 526 CTGCGATATTCCTGATGAGCGCTTTGACGCCACATTCACACTACGTTGGTGAATTC 585  
|||  
Db 273 CTAAACAGTGCCTATGAGCGCTTTGAGCCACATTCACACTACGTTGGTGAATTC 332  
|||  
QY 586 TTCTGGGCGATTGCCAGTACCTGCTCCAGGCATATTCAGAGTCTGCTACATCG 641  
|||  
Db 333 TTCTGGGCGATTGCCAGTACCTGCTCCAGGCATATTCAGAGTCTGCTACATCG 388

LOCUS 1555 bp mRNA PRI 19-DEC-1995  
DEFINITION Human alpha 7 neuronal nicotinic acetylcholine receptor mRNA, complete cds.  
ACCESSION U04583  
NID 91125076  
VERSION 91125076  
KEYWORDS alpha 7  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1977)  
AUTHORS Logel, J., Dreling, C., Barnhart, M., Antle, C. and Leonard, S.  
TITLE Nucleotide Sequence and Transcript Size of the Alpha-7 Neuronal Nicotinic Acetylcholine Receptor in Human Postmortem Brain  
JOURNAL Unpublished  
AUTHORS Leonard, S.  
TITLE Direct Submission  
JOURNAL Submitted (13-NOV-1995) Sherry Leonard, University of Colorado Health Sciences Center, C-268-71 Pharmacology, 4200 E. Ninth Ave, Denver, CO 80262, USA

FEATURES  
source  
1. 1977  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="pBSalpha7"  
/clone\_lib="Clontech lambda gt10 cDNA library: human hippocampus"  
/tissue\_type="hippocampus"  
/sex="male"  
/dev\_stage="20 yr old adult"  
/protein\_id="AAA83561.1"  
/db\_xref="PIR:G1125077"  
/db\_xref="GI:1125077"  
/translation="WRCSFGVWALASLHVSIOGEFQRLKELVKNPLERPV  
ANDSOPLVFSLNLIQIMVSEYPGVKTIVRPDGOIMKPDILLYNSADERPAT  
FHTNVLVNSGHCQYLPQIFKSCYIDRWPFQVQHCRIKGSWSYSGMSLDLQMO  
EADISGYNIGENDLVIGIKRSERYECCKEPYPTVTYRRRTLYGLNLIPIC  
VLISALAVFLPADSGERTIGITVLISLVFMILVAIMPATSDVPLIAQFAS  
TMTISGLSVVTVIVLQHHHDPDGMKMTFVILLNMCAMFLRMKRGEDKVRPAC  
OHKORSSLASVMSAVAPPASNGNLLYIGFGDGVHCVPTPDGGVCGAMPAPT  
HDEHLHGQPPGEGDPLAKILEEVRVIANRFRCQDESAVCSSEMKFACVVDRLCLM  
AFSEVFTICTIGILMSAPNFEAVSKDFA"

5'UTR  
sig\_peptide  
CDS

3'UTR  
polyA\_site  
BASE COUNT 426 a 567 c 524 g 460 t  
ORIGIN

